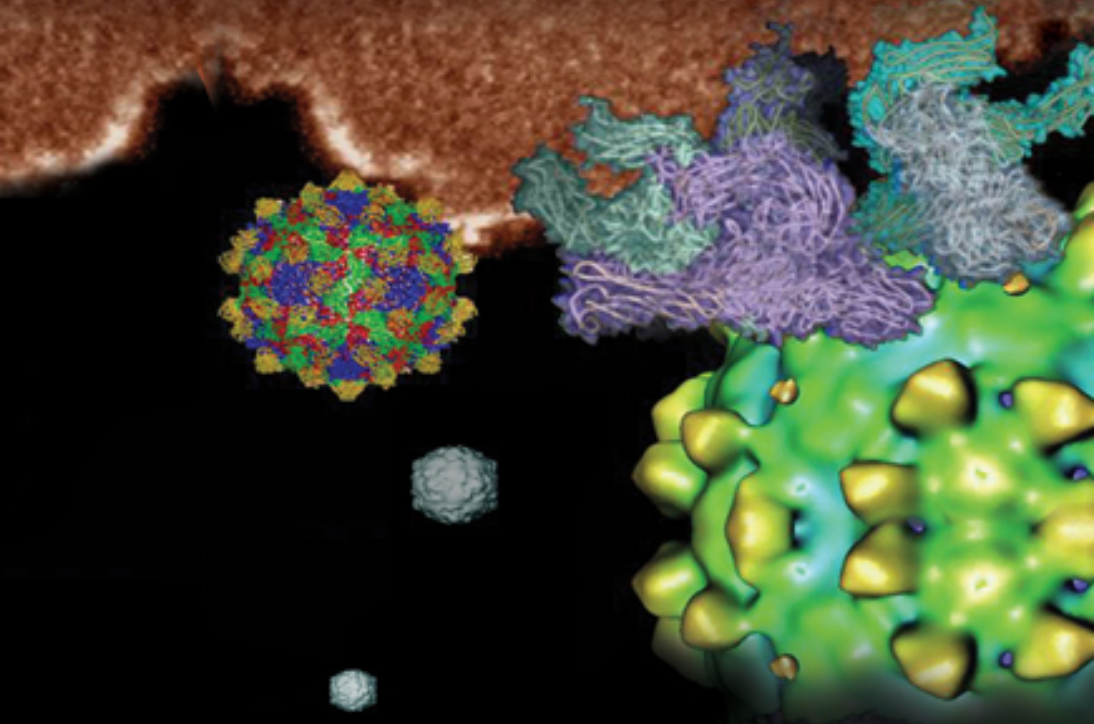
R. Holland Cheng · Tatsuo Miyamura

editors



Structure-based Study of VIRAL REPLICATION

(with CD-Rom)

World Scientific

VISIT…

Structure-based Study of VIRAL REPLICATION

(with CD-Rom)

This page intentionally left blank

Structure-based Study of VIRAL REPLICATION

(with CD-Rom)

Editors

R Holland Cheng

University of California, Davis, USA

Tatsuo Miyamura

The National Institute of Infectious Disease, Japan

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Structure-based study of viral replication : with CD-ROM / editors, R. Holland Cheng, Tatsuo Miyamura.

p. ; cm.

Includes bibliographical references and index.

ISBN-13: 978-981-270-405-4 (hardcover : alk. paper)

ISBN-10: 981-270-405-1 (hardcover : alk. paper)

ISBN-13: 978-981-270-406-1 (pbk. : alk. paper)

ISBN-10: 981-270-406-X (pbk. : alk. paper)

1. Viruses--Reproduction. 2. Molecular structure. I. Cheng, R. Holland. II. Miyamura, Tatsuo.

[DNLM: 1. Virus Replication. 2. Cellular Structures. 3. Morphogenesis. QW 160 S9273 2008]

QR470.S78 2008

579.2--dc22

2007038224

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Copyright © 2008 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

Typeset by Stallion Press

Email: enquiries@stallionpress.com

Printed in Singapore.

Contents

List of Contributors ix

Chapter 1 Human Rhinovirus Cell Entry and Uncoating 1
Renate Fuchs and Dieter Blaas

Chapter 2 Role of Lipid Microdomains in Influenza Virus Multiplication 43
Makoto Takeda

Chapter 3 Functions of Integrin $\alpha 2\beta 1$, A Collagen Receptor, in the Internalization of Echovirus 1 63
Varpu Marjomäki, Mikko Huhtala, Mikko Karjalainen, Timo Hyypiä and Jyrki Heino

Chapter 4 Entry Mechanism of Murine and SARS Coronaviruses — Similarity and Dissimilarity 77
Fumihiro Taguchi

Chapter 5 Hepatitis Viruses, Signaling Events and Modulation of the Innate Host Response 91
Syed Mohammad Moin, Anindita Kar-Roy and Shahid Jameel

Chapter 6 Virus-Cell Interaction of HCV 125
Hideki Tani, Yasumasa Komoda, Chang-kwang Limn, Kensuke Suzuki, Kohji Moriishi, Tatsuo Miyamura and Yoshiharu Matsuura

Chapter 7	RNA Replication of Hepatitis C Virus *Hideki Aizaki and Tetsuro Suzuki*	151
Chapter 8	Structure and Dynamics in Viral RNA Packaging *Thorsten Dieckmann and Marta Zumwalt*	173
Chapter 9	Rational Design of Viral Protein Structures with Predetermined Immunological Properties *James Lara and Yury Khudyakov*	199
Chapter 10	Bioinformatics Resources for the Study of Viruses at the Virginia Bioinformatics Institute *Anjan Purkayastha, Oswald Crasta, J. Dana Eckart, Michael Czar, X. J. Meng, Joao Setubal and Bruno Sobral*	267
Chapter 11	Virus Architecture Probed by Atomic Force Microscopy *A. J. Malkin, Yu. G. Kuznetsov, M. Plomp and A. McPherson*	289
Chapter 12	Filovirus Assembly and Budding *Takeshi Noda and Yoshihiro Kawaoka*	311
Chapter 13	Challenges in Designing HIV Env Immunogens for Developing a Vaccine *Indresh K. Srivastava and R. Holland Cheng*	327
Chapter 14	Insights into the *Caliciviridae* Family *Grant Hansman*	381
Chapter 15	Mathematical Approaches for Stoichiometric Quantification in Studies of Viral Assembly and DNA Packaging *Peixuan Guo, Jeremy Hall and Tae Jin Lee*	401

Chapter 16 Virus-like Particles of Fish Nodavirus 449
Chan-Shing Lin

Chapter 17 The Assembly of the Double-Layered Capsids of Phytoreoviruses 463
Toshihiro Omura, Naoyuki Miyazaki, Hisashi Naitow, R. Holland Cheng, Tomitake Tsukihara and Atsushi Nakagawa

Chapter 18 Structure and Assembly of Human Herpesviruses: New Insights From Cryo-Electron Microscopy and Tomography 483
Z. Hong Zhou and Pierrette Lo

Chapter 19 Human Papillomavirus Type 16 Capsid Proteins: Immunogenicity and Possible Use as Prophylactic Vaccine Antigens 517
Tadahito Kanda, Kei Kawana and Hiroyuki Yoshikawa

Chapter 20 Chimeric Recombinant Hepatitis E Virus-like Particles Presenting Foreign Epitopes as a Novel Vector of Vaccine by Oral Administration 539
Yasuhiro Yasutomi

Chapter 21 Nucleocapsid Protein of Hantaviruses (*Bunyaviridae*): Structure and Functions 553
Alexander Plyusnin, Vibhor Kumar, Olli Vapalahti and Peter Engelhardt

Chapter 22 Astrovirus Replication: An Overview 571
Susana Guix, Albert Bosch and Rosa M. Pintó

Chapter 23 DNA Vaccines against Viruses 597
Britta Wahren and Margaret Liu

Chapter 24 Life Cycles of Polyomaviridae — DNA Tumor Virus 609
Masaaki Kawano, Hiroshi Handa and R. Holland Cheng

Index 631

List of Contributors

Hideki Aizaki
Department of Virology II
National Institute of Infectious Diseases
1-23-1 Toyama
Shinjuku-ku
Tokyo 162-8640
Japan

Dieter Blaas
Max F. Perutz Laboratories
University Departments at the Vienna Biocenter
Department of Medical Biochemistry
Medical University of Vienna
Dr. Bohr Gasse 9/3
A-1030 Vienna
Austria

Albert Bosch
Department of Microbiology
University of Barcelona
Avda Diagonal 645
08028 Barcelona
Spain
and
Enteric Virus Group
Department of Microbiology
University of Barcelona

R. Holland Cheng
Professor of Molecular and Cellular Biology
Advanced Microscopy and Proteomics
University of California
Briggs Hall
Davis, CA 95616
USA
and
Karolinska Institute
Structural Virology
Novum, Hälsovägen 7
Huddinge, SE 14157
Sweden
and
Professor of Molecular and Cellular Biology
College of Biological Sciences
University of California
Davis Briggs Hall
Davis, CA 95616
USA

Oswald Crasta
Virginia Bioinoformatics Institute
Virginia Polytechnic and State University
Blacksburg, VA
USA

Michael Czar
Virginia Bioinoformatics Institute
Virginia Polytechnic and State University
Blacksburg, VA
USA

Thorsten Dieckmann
Deparment of Chemistry
University of Waterloo
200 University Avenue West
Waterloo
ON N2L 3G1
Canada

J. Dana Eckart
AFLAC
Columbus, GA
USA

Peter Engelhardt
Laboratory of Computational Engineering
Helsinki University of Technology
P.O. Box 9203 (Innopolis 2, Tekniikantie 14),
FI-02015 Espoo
Finland
and
Department of Pathology and Virology
Haartman Institute
University of Helsinki
P.O. Box 21 (Haartmaninkatu 3)
FI-00014
Helsinki

Renate Fuchs
Department of Pathophysiology
Medical University of Vienna
Währinger Gürtel 18-20
A-1090 Vienna
Austria

Susana Guix
Enteric Virus Group
Department of Microbiology
University of Barcelona

Peixuan Guo
1438 Vontz Center for Molecular Studies
University of Cincinnati
Cincinnati, OH 45267
USA
and
Department of Biomedical Engineering
College of Engineering and College of Medicine
University of Cincinnati
OH 45267
USA

Jeremy Hall
Department of Biomedical Engineering
College of Engineering and College of Medicine
University of Cincinnati
OH 45267
USA

Hiroshi Handa
Tokyo Institute of Technology
4259 Nagatsuta-cho
Midori-ku
Yokohama 226-8503
Japan

Grant Hansman
National Institute of Infectious Diseases
Tokyo

Jyrki Heino
Department of Biochemistry and Food Chemistry
University of Turku
Vatselankatu 2
FI-20014 Turku
Finland

Mikko Huhtala
Department of Biochemistry
Åbo Akademi University
Tykistökatu 6A
FI-20520 Turku
Finland

Timo Hyypiä
Department Virology
University of Turku
Kiinamyllynkatu 13
FI-20520 Turku
Finland

Shahid Jameel
Virology Group
International Centre for Genetic Engineering and Biotechnology
Aruna Asaf Ali Marg
New Delhi 110067
India

Tadahito Kanda
Division of Molecular Genetics
National Institute of Infectious Diseases
1-23-1 Toyama
Shinjuku-ku
Tokyo 162-8640
Japan

Mikko Karjalainen
Department of Biological and Environmental Sciences
University of Jyväskylä
Survontie 9
FI-40500 Jyväskylä
Finland

Anindita Kar-Roy
Virology Group
International Centre for Genetic Engineering and Biotechnology
Aruna Asaf Ali Marg
New Delhi 110067
India

Kei Kawana
Department of Obstetrics and Gynecology
University of Tokyo
7-3-1 Hongo
Bunkyo-ku
Tokyo 113-8655
Japan

Masaaki Kawano
Molecular and Cellular Biology
University of California
Davis CA 95616
USA

Yoshihiro Kawaoka
International Research Center for Infectious Diseases
Division of Virology
Department of Microbiology and Immunology
Institute of Medical Science
University of Tokyo
Shirokanedai
Minato-ku
Tokyo 108-8639

and
Core Research for Evolutional Science and Technology
Japan Science and Technology Agency
Saitama 332-0012
and
Department of Pathological Sciences
School of Veterinary Medicine
University of Wisconsin–Madison
Madison, Wisconsin 53706

Yury Khudyakov
Division of Viral Hepatitis
Centers for Disease Control and Prevention
1600 Clifton Road MS-A33
Atlanta, Georgia, 30333
USA

Yasumasa Komoda
Research Institute for Microbial Diseases
Osaka University
Osaka 565-0821
Japan
and
Department of Virology II
National Institute of Infectious Diseases
Tokyo 162-8640
Japan

Vibhor Kumar
Laboratory of Computational Engineering
Helsinki University of Technology
P.O. Box 9203 (Innopolis 2, Tekniikantie 14)
FI-02015 Espoo
Finland

Yu. G. Kuznetsov
Department of Molecular Biology and Biochemistry
University of California
Irvine, California 92697-3900
USA

James Lara
Division of Viral Hepatitis
Centers for Disease Control and Prevention
1600 Clifton Road MS-A33
Atlanta, Georgia, 30333
USA

Tae Jin Lee
Department of Biomedical Engineering
College of Engineering and College of Medicine
University of Cincinnati
OH 45267
USA

Chang-kwang Limn
Research Institute for Microbial Diseases
Osaka University
Osaka 565-0821
Japan
and
Department of Virology II
National Institute of Infectious Diseases
Tokyo 162-8640
Japan

Chan-Shing Lin
The Department of Marine Biotechnology and Resources
National Sun Yat-sen University
Kaohsiung 804
Taiwan

Margaret Liu
Department of Microbiology and Tumor Biology
Karolinska Institute
Swedish Institute for Infectious Disease Control
Stockholm
Sweden
and
Pro Ther Immune
Lafayette, California
USA

Pierrette Lo
Texas Heart Institute
PO Box 20345
MC3-116
Houston, TX 77225
USA

A. J. Malkin
Chemistry, Materials, and Life Sciences Directorate
Lawrence Livermore National Laboratory
Livermore, CA 94551
USA

Varpu Marjomäki
Department of Biological and Environmental Sciences
University of Jyväskylä
Survontie 9
FI-40500 Jyväskylä
Finland

Yoshiharu Matsuura
Research Institute for Microbial Diseases
Osaka University
Osaka 565-0821
Japan

and
Department of Virology II
National Institute of Infectious Diseases
Tokyo 162-8640
Japan

A. McPherson
Department of Molecular Biology and Biochemistry
University of California
Irvine, California 92697-3900
USA

X. J. Meng
VA-MD Regional College of Veterinary Medicine
Virginia Polytechnic and State University
Blacksburg, VA
USA

Tatsuo Miyamura
Research Institute for Microbial Diseases
Osaka University
Osaka 565-0821
Japan
and
Department of Virology II
National Institute of Infectious Diseases
Tokyo 162-8640
Japan

Naoyuki Miyazaki
Institute for Protein Research
Osaka University
3-2 Yamadaoka, Suita
Osaka 565-0871
Japan

Syed Mohammad Moin
Virology Group
International Centre for Genetic Engineering and Biotechnology
Aruna Asaf Ali Marg
New Delhi 110067
India

Kohji Moriishi
From the Research Institute for Microbial Diseases
Osaka University
Osaka 565-0821
Japan
and
Department of Virology II
National Institute of Infectious Diseases
Tokyo 162-8640
Japan

Hisashi Naitow
Division of Bio-Crystallography Technology
RIKEN Harima Institute
1-1-1 Kouto Mikazuki
Sayo
Hyogo 679-5148
Japan

Atsushi Nakagawa
Institute for Protein Research
Osaka University
3-2 Yamadaoka, Suita
Osaka 565-0871
Japan

Takeshi Noda
International Research Center for Infectious Diseases
Core Research for Evolutional Science and Technology
Japan Science and Technology Agency
Saitama 332-0012

Toshihiro Omura
National Agricultural Research Center
3-1-1 Kannondai
Tsukuba
Ibaraki 305-8666
Japan

Rosa M. Pintó
Enteric Virus Group
Department of Microbiology
University of Barcelona

M. Plomp
Chemistry, Materials, and Life Sciences Directorate
Lawrence Livermore National Laboratory
Livermore, CA 94551
USA

Alexander Plyusnin
Department of Virology
Haartman Institute
University of Helsinki
P.O. Box 21 (Haartmaninkatu 3)
FI-00014 Helsinki
Finland

Anjan Purkayastha
Virginia Bioinoformatics Institute
Virginia Polytechnic and State University
Blacksburg, VA
USA

Joao Setubal
Virginia Bioinoformatics Institute
Virginia Polytechnic and State University
Blacksburg, VA
USA

Bruno Sobral
Virginia Bioinoformatics Institute
Virginia Polytechnic and State University
Blacksburg, VA
USA

Indresh K. Srivastava
Associate Director
Vaccines Research
Novartis Vaccines and Diagnostics, Inc.
Mail Stop 4.3, 4560 Horton Street
Emeryville, CA 94608

Kensuke Suzuki
Research Institute for Microbial Diseases
Osaka University
Osaka 565-0821
Japan
and
Department of Virology II
National Institute of Infectious Diseases
Tokyo 162-8640
Japan

Tetsuro Suzuki
Department of Virology II
National Institute of Infectious Diseases
1-23-1 Toyama
Shinjuku-ku
Tokyo 162-8640
Japan

Fumihiro Taguchi
Laboratory of Respiratory Viral Diseases and SARS
National Institute of Infectious Diseases
Murayama Branch
4-7-1 Gakuen
Musashi-Murayama
Tokyo 208-0011
Japan

Makoto Takeda
Department of Virology
Faculty of Medicine
Kyushu University
Fukuoka 812-8582
Japan

Hideki Tani
Research Institute for Microbial Diseases
Osaka University
Osaka 565-0821
Japan
and
Department of Virology II
National Institute of Infectious Diseases
Tokyo 162-8640
Japan

Tomitake Tsukihara
Institute for Protein Research
Osaka University
3-2 Yamadaoka, Suita
Osaka 565-0871
Japan

Olli Vapalahti
Department of Virology
Haartman Institute
University of Helsinki
P.O. Box 21 (Haartmaninkatu 3)
and
Department of Basic Veterinary Sciences
Faculty of Veterinary Medicine
P.O. Box 66
FIN-00014
University of Helsinki and HUSLAB
Helsinki
Finland

Britta Wahren
Department of Microbiology and Tumor Biology
Karolinska Institute
Swedish Institute for Infectious Disease Control
Stockholm
Sweden

Yasuhiro Yasutomi
Laboratory of Immunoregulation and Vaccine Research
Tsukuba Primate Research Center
National Institute of Biomedical Innovation
Tsukuba
Ibaraki 305-0843
Japan
and
Department of Immunoregulation
Mie University Graduate School of Medicine
Mie 514-8507
Japan

Hiroyuki Yoshikawa
Department of Obstetrics and Gynecology
University of Tsukuba
1-1-1 Tennoudai
Tsukuba-shi
Ibaraki 305-8575
Japan

Z. Hong Zhou
Department of Pathology and Laboratory Medicine
University of Texas Medical School at Houston
Houston, TX 77030
USA
and
Department of Microbiology
Immunology and Molecular Genetics
David Geffen School of Medicine, and
the California NanoSystems Institute
University of California at Los Angeles
Los Angeles, CA 90095
USA

Marta Zumwalt
Deparment of Chemistry
University of Waterloo
200 University Avenue West
Waterloo
ON N2L 3G1
Canada

Chapter 1

Human Rhinovirus Cell Entry and Uncoating

Renate Fuchs[*,†] *and Dieter Blaas*[‡]

Human rhinoviruses (HRVs) are small, icosahedral, non-enveloped, single-stranded positive-sense RNA viruses. Out of the 74 type A serotypes 12, the minor group bind members of the LDL receptor-family; the remainder plus all the 25 type B HRVs bind intercellular adhesion molecule-1. HRVs enter cells by receptor-mediated endocytosis. The ensuing structural modifications lead to release of the viral RNA into the cytosol where virus replication takes place. Binding to plasma membrane receptors, entry into the cell, uncoating, and penetration of the viral genome are discussed with respect to receptor and virus structure. Despite high structural similarity, major and minor group HRVs, as well as the individual major group serotypes, differ with respect to the process of entry and uncoating.

Introduction

Since the isolation of a common cold virus from nasal mucus and its propagation in tissue culture,[1] much has been learned about the replication cycle of these small, icosahedral, single-stranded positive-sense

*Corresponding author.

[†]Department of Pathophysiology, Medical University of Vienna, Währinger Gürtel 18-20, A-1090 Vienna, Austria; E-mail: renate.fuchs@meduniwien.ac.at.

[‡]Max F. Perutz Laboratories, University Departments at the Vienna Biocenter, Department of Medical Biochemistry, Medical University of Vienna, Dr. Bohr Gasse 9/3, A-1030 Vienna, Austria; E-mail: dieter.blaas@meduniwien.ac.at.

RNA viruses termed "human rhinoviruses" (HRVs). Much knowledge stems from earlier studies of the related enteroviruses, in particular the polio virus and coxsackievirus, which are rather closely related to HRVs; with due care, insight gained from work on enteroviruses can often, but not always, be extrapolated to the rhinovirus field. Physicochemically, rhinoviruses are distinguished from enteroviruses based on their acid lability, a feature originally used for their classification; enteroviruses, in contrast, remain infective at pH below 3; thus, they can pass unharmed through the digestive tract. They infect the intestinal epithelia and sometimes spread throughout the body, causing viremia. Conversely, rhinoviruses are comparably harmless, usually remaining confined to the upper respiratory tract and only occasionally spreading to the lungs.

During the HRV infection cycle, the following sequence of events can be differentiated: (1) virus binding to its receptors at the plasma membrane; (2) entry into the cell by receptor-mediated endocytosis; (3) conformational change of the viral capsid; (4) release of the viral RNA ("uncoating"); (5) RNA penetration into the cytoplasm; (6) synthesis of viral proteins; (7) RNA replication; and (8) assembly and release of new, infectious virions.

HRVs are composed of a protein shell assembled from 60 copies each of the four capsid proteins VP1, 2, 3, and 4. VP4 is internal and in close proximity to the RNA; however, due to the dynamic nature of the capsid, large parts of VP4 and the capsid-internal N-terminus of VP1 become temporally exposed to the solvent, a feature termed "breathing."[2–6] The viral shell is about 30 nm in diameter with the five-fold axes of icosahedral symmetry being surrounded by a cleft, termed the canyon. It encloses a single-stranded RNA genome of roughly 7100 bases. Upon arrival in the cytosol, the RNA becomes translated into a polyprotein that is autocatalytically and co-translationally cleaved by the viral proteinases $2A^{pro}$, $3C^{pro}$ and $3CD^{pro}$ into VP1, VP0, VP3 and the non-structural proteins. Maturation cleavage of VP0 into VP2 and VP4 occurs by an unknown protease upon virus assembly. Not counting the precursor proteins — such as 3CD, the precursor of the protease $3C^{pro}$ and the RNA-dependent RNA polymerase $3D^{pol}$ — 11 mature polypeptides are eventually generated from the polyprotein.

The large number of serologically different rhinovirus types has precluded the development of vaccines, and antiviral agents have still not reached clinical application. The latter are targeted to the viral capsid and prevent uncoating by binding into a hydrophobic pocket or to virally encoded enzymes, such as the proteases.

Rhinovirus Receptor Groups and Subgenera

Early on, Lonberg-Holm and colleagues noted that some human rhinovirus serotypes competed for binding to the cell surface, suggesting that they use the same cellular receptors.[7,8] Based on a more extensive analysis using blockage by a monoclonal antibody, 91 serotypes were assigned to the major receptor group and 10 serotypes to the minor receptor group.[9–11] The major and the minor group receptors were subsequently shown to be identical to the intercellular adhesion molecule 1 (ICAM-1)[12–14] and members of the low-density lipoprotein receptor (LDLR) superfamily, respectively. The latter includes LDLR, the very-LDLR (VLDLR) and the LDLR related protein (LRP).[15–18] HRV87 could not be allocated to either group, and, based on phylogeny, it was later found to be a prime strain of the acid-labile enterovirus EV68.[19–21] This classification has been generally adopted by the scientific community. However, recently HRV23 and HRV25 were found to be minor group viruses.[22] Since serotype Hanks and HRV21, as well as HRV8 and HRV95, exhibit extensive immunologic cross-reactivity and high amino acid sequence similarity of VP1,[20,23] they can be considered as two single serotypes. Therefore, there are now 87 major group HRVs and 12 minor group HRVs (HRV1A and HRV1B being regarded as two separate strains).

Another means of HRV classification emerged from the observation that different rhinovirus serotypes were differently sensitive towards a panel of antivirally active compounds; many of these are derived from isoxazol, and were termed "WIN-compounds" since they were synthesized by the Sterling Winthrop company.[24] Whereas some HRVs are neutralized by compounds with a long aliphatic chain, others preferentially bind shorter compounds[25]; this gave rise to cataloguing HRVs as antiviral groups A and B.[26] No correlation between these antiviral groups and the two receptor groups was found. Nevertheless,

binding of these drugs is clearly related to the geometry of the hydrophobic pocket in VP1, which lies beneath the canyon floor. In several HRV serotypes, this pocket is filled with a natural pocket factor, presumably a fatty acid that becomes displaced by the antiviral because of its higher affinity.[27] Several studies have shown that the drugs act by stabilizing the capsid against structural modifications that are required for RNA release.[3,4,28–30] In some cases, particularly in HRV14, the canyon floor is also shifted upwards upon binding of the drugs and the receptor attachment site becomes modified, resulting in loss of affinity for ICAM-1 (see below).[31] Recently, the nucleotide sequence of VP1 of all rhinovirus serotypes has become available, and on the basis of phylogeny the 99 serotypes were divided into 74 type A and 25 type B viruses.[20,21,23,32] However, the similarity within serotypes of the same receptor group was not superior to that within serotypes belonging to different groups. The basis of receptor discrimination thus remains unclear. Interestingly, the genetic classification rather closely follows the classification based on sensitivity toward the antivirals. This underscores the correlation between amino acid sequence and geometry of the pocket. Unfortunately, the terminology used is somewhat confusing; the majority of the members of antiviral group A belong to genus B and the majority of antiviral group B corresponds to genus A.

Virus Structure

Rhinoviruses are T = 1, pseudo T = 3 icosahedrons since the three external capsid proteins adopt a very similar β-barrel conformation to that seen in the T = 3 plant viruses built from 180 identical copies of a single capsid protein. The longer loops connecting the antiparallel β-sheets are exposed to the solvent; the shorter ones are internal and close to the RNA. The serotypic variability stems from the different amino acid sequences of the external loops that constitute the antigenic epitopes.[33,34] In 1989, Michael Rossmann predicted that the canyon, a cleft encircling each of the 12 vertices of the icosahedron, would be involved in receptor recognition and would accommodate five receptor molecules around each of the 12 symmetry axes. According to this "canyon hypothesis," residues at the bottom of the canyon would be conserved within the

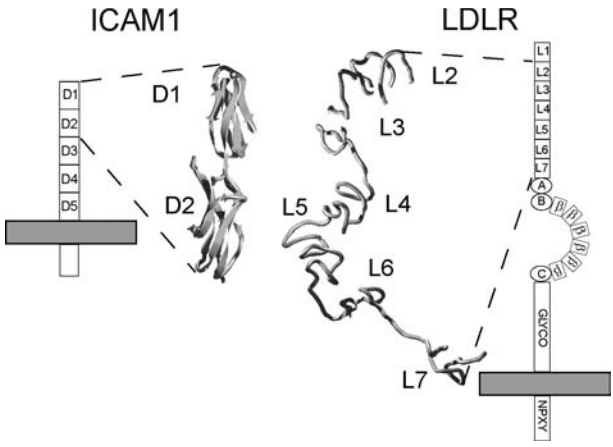


Fig. 1. Comparison of the two receptors used by human rhinoviruses for cell entry. The structure of the first two immunoglobulin domains of the major group receptor ICAM-1[43] and of six ligand binding repeats (L2 through L7; L1 is disordered in the structure and not seen) of LDLR at pH 5.3[44] are shown. The drawings schematize the orientation of the receptors with respect to the plasma membrane (grey rectangle) and include other domains present in the native receptors. A, B, and C denote domains with similarity to epidermal growth factor precursor. Empty rectangles symbolize the YWTD domains forming a β-propeller. Note the absence of a coated pit localization signal (NPXY) in ICAM-1.

different serotypes since the geometry of the cleft would prevent entry of antibodies and thus immunological pressure; at the same time, the cleft would be accessible for a slim receptor molecule.[35] This proved to be true in part since the receptor used by the major group of HRVs, the slender ICAM-1 molecule (Fig. 1), indeed binds within the canyon[36] and the residues in this region appear to be somewhat more conserved than the loop residues.[37] Alignment of VP1 sequences of all major group HRVs with respect to non-contiguous amino acid residues known to contact ICAM-1 in HRV3, 14 and 16, revealed two patterns strictly conserved within[37a] the two subgenera. However, antibodies can also penetrate rather deeply into this cleft.[38]

The canyon hypothesis was further challenged by the discovery that, in the minor group HRVs, the receptors do not bind within the

canyon but rather attach to the most exposed star-shaped dome that is encircled by the canyon. Three-dimensional structures of the rhinovirus serotypes HRV1A, HRV2, HRV3, HRV14, and HRV16 are available at atomic resolution,[27,39–42] but the principles of receptor discrimination could not be elucidated from these data.

The Receptors

ICAM-1 and LDL-receptors, both glycosylated type I transmembrane proteins, are completely different from each other in terms of both function and structure (Fig. 1). ICAM-1 is predominantly involved in cell-cell adhesion and in immune reactions by binding to the integrins lymphocyte function associated antigen 1 (LFA1) and macrophage differentiation antigen (Mac-1). Its extracellular part has a rod-like structure of about 190 Å in length[45] and is composed of five typical immunoglobulin-like domains of some 60 amino acid residues each.[46] It possesses a short cytoplasmic tail of 29 residues without typical clathrin-coated pit localization signals. An X-ray structure of the two N-terminal domains at 2.2 Å is available.[47] Each domain is roughly 35 Å long and has a diameter of about 20 Å.[43] Conversely, members of the LDLR family function in ligand internalization and signal transduction. Their ligand-binding domains are composed of different numbers of ligand-binding repeats, or modules, containing about 40 amino acid residues each. LDLR has seven such repeats, while VLDLR has eight and LRP has 31. In LRP these are arranged in clusters of 2, 8, 10, and 11 modules, counting from the N-terminus. The repeats exhibit an ellipsoidal shape with axes of approximately 23 Å and 20 Å in length.[48] In LDLR and VLDLR, the ligand-binding domain is followed by three regions with similarity to the epidermal growth factor precursor (EGF-domain) that contain YWTD motives forming a six-bladed β-propeller, a domain with O-glycosylation proximal to the membrane, a transmembrane domain, and a cytoplasmic tail with an NPXY internalization motive.[49,50] The other members of the LDLR family exhibit similar domain arrangements; LRP might be considered a mosaic of several LDLR fragments with interspaced EGF-domains and a longer cytoplasmic extension with several internalization motives.[51]

Virus-receptor Complexes

A number of picornaviruses in complex with their respective recombinant soluble receptor fragments have been analyzed by cryo-electron microscopy (cyro-EM).[36,52–56] These medium-resolution structures, together with available X-ray structures of the viruses and receptors, were used for building models of the complexes that have, at present, attained a resolution of less than 10 Å.[57] Although, this does not allow for the detailed interpretation of amino acid residue contacts, it permits rather exact delineation of the interface. For example, for the complexes between HRV14 or HRV16 and ICAM-1 the footprint area was determined to be about 1,400 Å^2; the parameters of HRV14 and HRV16 binding to wt ICAM-1 and to an ICAM-1 mutant were interpreted in the context of the contact residues.[57]

The first X-ray structure of a complex between the minor group rhinovirus HRV2 and two modules of the very-low density lipoprotein receptor (V23) has been solved and revealed an intricate pattern of ionic, electrostatic and hydrophobic interactions.[58] In this case, the surface area of the receptor footprint is only 430 Å^2, and is therefore much inferior to that of the complexes between the major group viruses and ICAM-1. This has important consequences for the binding mode of these receptors.

In the following section we will concentrate on the early events of infection of representatives of the two receptor groups of human rhinoviruses, exemplified by the most extensively studied major group virus HRV14 (and in some cases HRV3 and HRV16) and the minor group virus HRV2. In Fig. 2 the different receptor binding sites on the respective viruses are illustrated in a side view of VP1 of HRV2 and HRV16. Whereas the tip of the amino terminal domain of ICAM-1 is seen to enter the canyon of HRV16, the single V3 module of VLDLR attaches sidewise to the protruding BC, DE, and HI loops of VP1 of HRV2. The superposition also demonstrates the very similar folding pattern of VP1 of the two viral serotypes. Figure 3 depicts the orientation of ICAM-1 versus V3 on a pentamer of the respective virus. Bound receptors were tilted by 180° to the right to allow viewing of the contact areas. Again, only the tip of the first domain of ICAM-1 is seen to

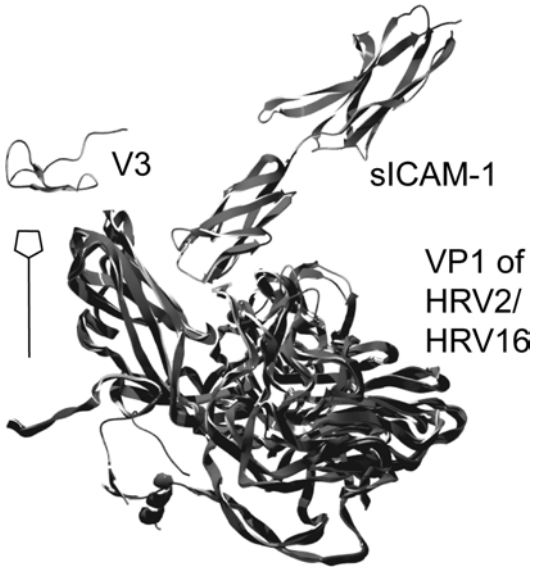


Fig. 2. Arrangement of V3 and ICAM-1 with respect to their cognate HRVs as seen in the complexes solved by cryo-EM (HRV16; PDB entry 1d3e) and X-ray (HRV2; PDB entry 1v9u). Side view onto a slightly inclined VP1 of HRV2 and HRV16 with the respective soluble recombinant receptor fragments attached. Coordinates of VP1 with attached receptor fragments from the cryo-EM structures of the complex between HRV16 and a two-domain ICAM-1 (PDB entry 1d3e) and of the X-ray structure of the complex between HRV2 and V23 (PDB entry 1v9u) were superimposed. The approximate position of the five-fold axis of icosahedral symmetry is indicated.

be in contact with the canyon, whereas the second domain sticks out into the solvent. In the cryo-EM structure of a complex between Coxsackievirus 21 (Cox21), an enterovirus also binding ICAM-1, and the entire exodomain of ICAM-1, all the five Ig-domains are visible, although the density of domains four and five is very low.[59] This indicates that ICAM-1 is extremely rigid. The model also gives an impression of the distance between the virus and the plasma membrane when attached to the host cell.[59] In contrast, in the X-ray structure of the complex between V23 and HRV2, the receptors are arranged around the five-fold axis resulting in a ring-like appearance

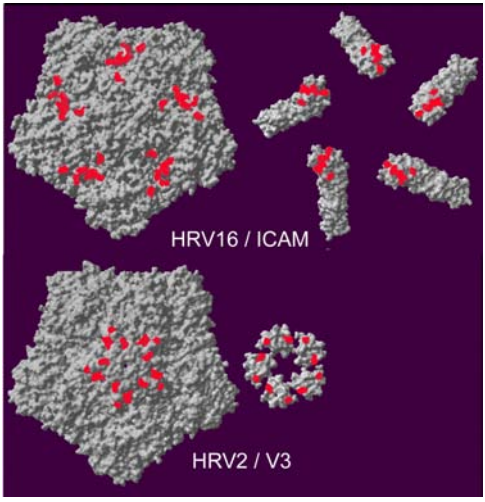


Fig. 3. Comparison of the binding sites of V3 and of ICAM-1 on the respective surfaces of HRV2 and HRV16. View onto a viral pentamer. The receptors were tilted by 180° to the right to visualize the respective contacts. Residues in HRVs and in the receptors within a distance of ≤ 3 Å from each other are colored red. Note that the five individual molecules of V3 come so close to each other that they appear to form a ring.

(Fig. 3). This could either result from the binding of five copies of V3 to the five symmetry-related sites, with each V2 moving freely, or from simultaneous binding of V2 and V3, contributed from the same molecule, to two neighboring symmetry-related sites, leaving one site free. This should give rise to a stoichiometry of 48 modules per virion (i.e. an occupancy of 80%), as was indeed found in the reconstructions.[58] Although the amino acid sequences of V2 and V3 differ in quite a number of positions, except those in contact with the virus that are at least functionally conserved, the fitting into the electron density was ambiguous; cryo-EM and X-ray crystallography are averaging techniques that do not easily allow differentiation between these possibilities. Therefore, under the conditions of crystal formation either 80% of the binding sites on the virus were occupied by V3 modules only, or each vertex was occupied by two copies of the concatemer V23.

The representation in Fig. 3 also demonstrates that the contact areas in the V23/HRV2 complex are much smaller than those of the ICAM-1/HRV16 complex. However, taking into consideration simultaneous binding of more than one or even five VLDLR modules contributed from the same protein, the contact area between the minor group receptor and HRV2 even exceeds that of the major group receptor and HRV16. Multi-module binding was definitely demonstrated by the analytical separation of complexes between HRV2 and an artificial pentamer of V3 fused to maltose binding protein (MBP-V33333). Capillary electrophoresis resolved peaks corresponding to virus carrying between zero and 12 receptor molecules, indicating that the pentamers attach to the 12 vertices of the virus, most probably involving all five modules. This view is also supported by the tremendous increase in avidity with the number of modules present in the recombinant molecule.[60,60a]

HRV Entry: Endocytic Pathways and Selective Inhibitors

For infection to occur, virus, subviral particles, or the viral genome has to enter the host cell. HRVs enter by receptor-mediated endocytosis and, more importantly, are dependent on endocytosis for infection (see below). Consequently, structural modifications of the viral capsid, as required for RNA release and penetration into the cytoplasm, occur from endocytic compartments. Since endosomal subcompartments maintain different internal environments, it is important to define the respective endosome population where these events take place. The mechanism of rhinovirus entry, uncoating, and infection has been studied predominantly in HeLa cells, which express ICAM-1 as well as LDLR, LRP, and VLDR; furthermore, in these cells almost all HRV serotypes replicate to high titers. Due to the difficulty of growing primary airway epithelial cells, the main site of HRV replication in humans, few studies have been carried out in this system. A main drawback of using HeLa cells is their lack of forming polarized monolayers on permeable filters despite their being derived from a cervix carcinoma; this precludes investigations of vectorial uptake or release of the virus.

After binding to plasma membrane receptors, virus uptake can occur by clathrin-mediated endocytosis, caveolae-mediated endocytosis, and/or clathrin and caveolae-independent internalization.[61,62] Clathrin-mediated uptake is dependent on particular sequence motives in the cytoplasmic domain of transmembrane receptors and can occur constitutively; in contrast, internalization of caveolae is triggered by e.g. clustering of receptors. Apart from specific coat proteins and accessory molecules, the GTPase dynamin plays an essential role in the constriction/fission process during clathrin- and caveolae-mediated uptake. Thus, over-expression of dominant-negative, non-functional mutant dynamin arrests ligand/receptor complexes at the plasma membrane.[61,62] The least characterized pathway is the clathrin and caveolae-independent internalization, which may involve dynamin and/or lipid rafts.[63] Irrespective of the uptake mechanism, internalized receptors and ligands are first delivered to early (sorting) endosomes.[64,65] However, there are cell-type specific variations with respect to the intersection of distinct endocytic pathways, the mechanism of sorting, and how these are affected by inhibitors. From early endosomes, internalized material can follow different intracellular routes[66]: (1) transport to lysosomes, resulting in degradation of ligands and certain receptors; (2) recycling to the cell surface; and (3) in polarized cells, transport from one plasma membrane domain to the opposite plasma membrane domain (transcytosis).

A major property of endocytic (and of some exocytic) compartments is the low internal pH, which is established and maintained by the activity of the vacuolar H^+-ATPase (V-ATPase).[67] The control of acidification and consequently the establishment of distinct pH values in endocytic sub-compartments play an important role in trafficking of macromolecules through endocytic pathways.[68,69] The mildly acidic pH in early endosomes causes the dissociation of many ligands from their receptors, allowing for receptor recycling, while the bulk of volume containing the released ligands is routed through late endosomes to lysosomes, where rapid degradation takes place (Fig. 4). In most cell types (e.g. HeLa cells), transfer from early to late endosomes involves endosomal carrier vesicles (ECV).[70,71] ECV and late endosomes contain many internal vesicles and are also referred to as multi-vesicular

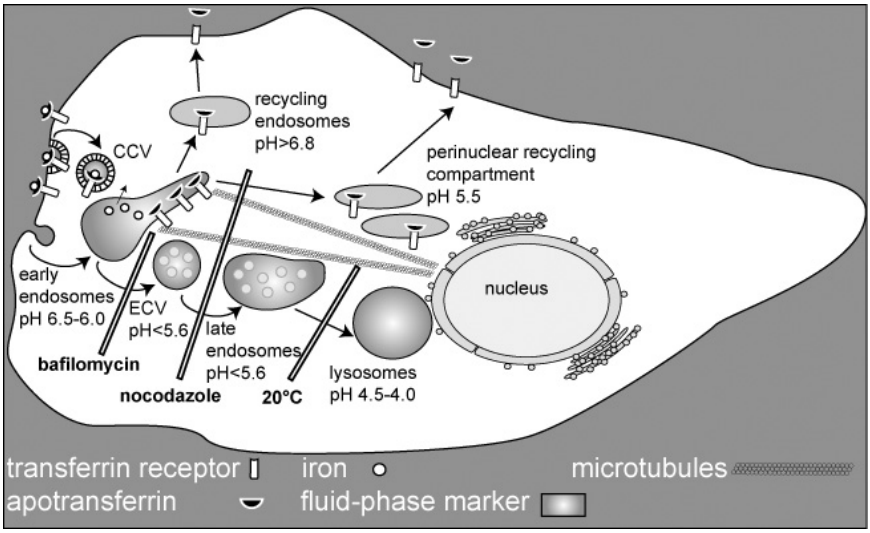


Fig. 4. Influence of bafilomycin, nocodazole and low temperature on endocytic pathways in HeLa cells. The transferrin recycling pathway and the transport of fluid-phase marker to lysosomes is shown. Iron-loaded transferrin binds to its receptor at the plasma membrane. The complex is internalized via clathrin-coated vesicles (CCV) and delivered into early endosomes within two to five minutes, where the iron is released and transferred into the cytoplasm. Apotransferrin remains bound to the receptor and recycles via a fast and slow pathway. At the plasma membrane, apotransferrin is released at the neutral pH. The main route of a fluid-phase marker (e.g. dextran) from early endosomes (within five min) via endosomal carrier vesicles (ECV) and late endosomes (within 15 min) to lysosomes (within 25 min) is shown in dark grey. Transferrin transport to and recycling via the PNRC is blocked by nocodazole, whereas bafilomycin and reduced temperature (20°C) are without effect. In contrast, bafilomycin arrests fluid-phase marker in early endosomes by preventing budding of ECV, whereas nocodazole leads to accumulation of cargo in ECV. Finally, incubation at 20°C prevents delivery of markers from late endosomes to lysosomes. The endosomal pH values shown are derived from FACS analysis of FITC/Cy5-dextran and FITC/Cy5-transferrin labeled compartments as well as from determination of the conformational change of HRV2 occurring at a threshold of pH ≤ 5.6.[70,73]

bodies. These internal vesicles contain receptor-ligand complexes that are destined for lysosomal degradation.[72] Material en route to lysosomes (pH 4.5–4.0) is exposed to an increasingly acidic pH during transport from early endosomes (pH 6.5–6.0) to EVC and late endosomes (pH ≤ 5.6).[66,68,73,74] Budding of ECV from early endosomes requires their acidification and the presence of distinct coat proteins.[71] Thus, inhibition of the V-ATPase by bafilomycin selectively blocks transport of cargo destined for lysosomal degradation in early endosomes. ECV are subsequently transported along microtubules to late endosomes. Disruption of microtubules by nocodazole results in accumulation of endocytosed material in ECV.[75] Finally, incubation of the cells at 20°C results in retention of cargo in late endosomes and thus prevents degradation in lysosomes.[76] Drugs and temperature can thus be used to arrest internalized material in different compartments on their way to lysosomes depending on the cell type (Fig. 4). These treatments also have distinct effects on the recycling pathway. This route is taken by (iron-loaded) transferrin and the transferrin receptor.[77] Therefore, the transferrin pathway is most often used for control purposes in investigations of virus entry pathways.

The transferrin-transferrin receptor complex is taken up via clathrin-coated pits and vesicles into early endosomes, where the iron is released.[77] The resulting apo-transferrin remains bound to the receptor and is recycled and subsequently released into the neutral extracellular milieu. Transferrin recycling can occur via two pathways: from early endosomes with $t_{1/2} \approx 2.5$ min (fast) and from the perinuclear recycling compartment (PNRC) $t_{1/2} \approx 7$ min (slow).[78,79] The pH of the PNRC is higher than that of early endosomes in CHO and Hep2 cells, whereas the PNRC is more acidic than early endosomes in HeLa cells.[73,77] Biochemically, the PNRC is distinct from early endosomes as well as from ECV and late endosomes. It contains recycling receptors such as LDLR, but no material that is targeted to lysosomes.[77] We have recently shown that the kinetics of transferrin acidification in HeLa cells is biphasic, indicative of fast and slow recycling pathways via early endosomes (pH 6.0) and the PNRC (pH 5.6). Nocodazole blocks the second phase of transferrin acidification, demonstrating that transferrin transport from

early endosomes to the PNRC is absolutely microtubule dependent.[73] In contrast, bafilomycin does not affect transferrin delivery to the PNRC and incubation at 20°C only delays transferrin recycling. Taken together, at least in HeLa cells, the transferrin recycling pathway and the transport of endocytosed material to lysosomes are differentially affected by bafilomycin, nocodazole and low temperature (Fig. 4). Consequently, these treatments can be applied to investigate whether viruses follow a recycling or degradative pathway and to identify the compartment where virus penetration/uncoating takes place.

Entry of Major Group HRVs

ICAM-1, the major group HRV receptor, lacks any known internalization signal in its cytoplasmic domain, and infection by HRV14 also occurs when the cytoplasmic tail of the receptor is removed or the transmembrane domain replaced by a GPI-anchor.[80] However, the virus appears to be internalized via clathrin-mediated endocytosis since overexpression of the non-functional dynamin 2 mutant K44A in HeLa cells resulted in inhibition of infection; involvement of caveolae is unlikely since these cells express little caveolin.[81] Given the absence of a *bona fide* internalization signal in ICAM-1, it has been speculated that a co-receptor may be involved. For example, adenovirus requires $\alpha_v\beta_3$ or β_5-integrins for internalization, although it first binds to the coxsackie-adenovirus receptor (CAR).[82] No evidence for an HRV co-receptor has been presented so far and *in vivo* data showing co-localization of major group virus with ICAM-1 in a particular compartment are still lacking. Although the affinity of recombinant ICAM-1 fragments for immobilized HRV3 decreases about 50-fold upon lowering of the pH from 8.0 to 6.0,[83] it has not been investigated *in vivo* whether the virus dissociates from ICAM-1 at low endosomal pH.

By using immunofluorescence microscopy, HRV14 was found in typical endosomal compartments (Fig. 5) when internalized at 20°C.[84] Under this condition the conformational modification of the capsid, catalyzed by ICAM-1, is inhibited (see below). However, when

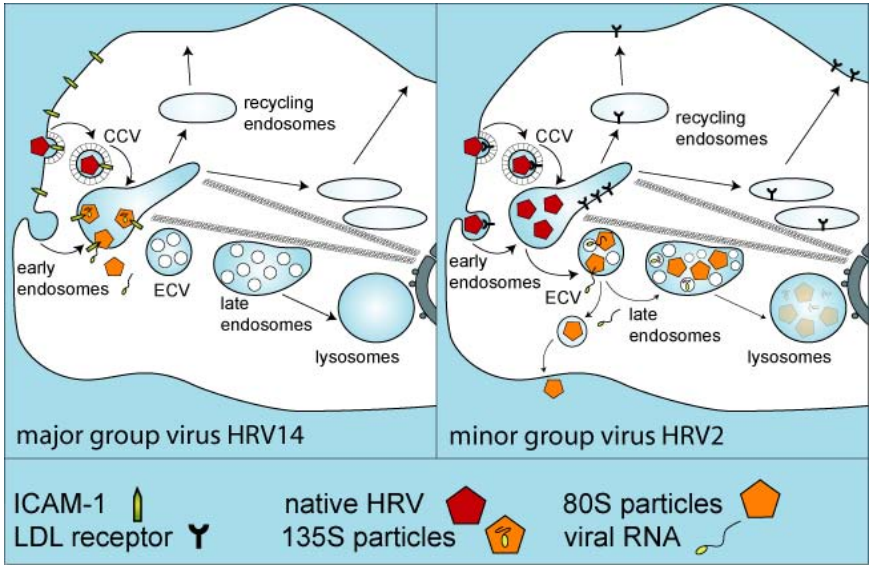


Fig. 5. Entry pathways of major and minor group HRVs in HeLa cells, exemplified by HRV14 and HRV2. HRV14 is internalized via clathrin-coated pits and vesicles (CCV) and delivered into early endosomes, where the receptor-catalyzed conformational change can take place at 34°C. Presumably, the structural modification is coupled to RNA release and rupture of the endosomal membrane resulting in delivery of free RNA and empty capsids into the cytoplasm. Consequently, the majority of virus escapes further transport to lysosomes. At least HRV3 and HRV14 can uncoat at neutral endosomal pH in the presence of bafilomycin and thus in early endosomes. HRV2 enters via clathrin-dependent and independent pathways and dissociates from its receptors at mildly acidic pH in early endosomes. Receptors are recycled and HRV2 is transferred to endosomal carrier vesicles (ECV), where the more acidic pH (≤ 5.6) induces the structural modification, resulting in RNA uncoating and transfer into the cytosol. Note that ECV and late endosomes are multivesicular bodies and thus contain many internal vesicles. Presumably, the RNA can also be translocated into the internal vesicles of ECV. Finally, residual native virus, subviral particles and viral RNA are transported via late endosomes to lysosomes, where they are degraded.

internalized at 34°C, the virus is apparently not transported to lysosomes as shown by the lack of viral RNA degradation products after 60 min.[85] This may indicate that HRV14 follows the recycling pathway, or that it is targeted to other organelles or escapes from the endosomes.

Entry of Minor Group HRVs

In contrast to ICAM-1, the receptors of minor group HRVs, LDLR, LRP and VLDLR, possess one or more clathrin-localization signals in their cytoplasmic tails.[86] Nevertheless, HRV2 can also be internalized by a clathrin-independent pathway when the clathrin-dependent route is inhibited by potassium depletion or by overexpression of non-functional dynamin K44A.[87–89] In the absence of such treatments, HRV2 may enter entirely via clathrin-coated pits and vesicles.[90] Similar to the natural ligands of LDLR,[49] HRV2 dissociates from its receptors at a pH of 6.0.[91] This conclusion is derived from experiments in which the pH-dependence of HRV2 dissociation from the plasma membrane of HeLa cells was analyzed at 4°C, a condition allowing virus binding to plasma membrane receptors, but not internalization. Furthermore, plasma membrane expression of LDLR and LRP is not altered during continuous HRV2 internalization (unpublished observations), indicative of receptor recycling. Since early endosomes maintain a mildly acidic pH of 6.0, which induces dissociation of HRV2 from its receptor(s), LDLR and LRP presumably recycle through the PNRC, whereas HRV2 becomes part of the fluid-phase of the endosomes and is sorted into ECV (Fig. 5). The virus is then transported via late endosomes to lysosomes, where it is degraded.[85,92]

Conversion of Native Virus into Subviral Particles: Role of Receptors and Low pH

Structural changes of the capsid occur within the endosomal compartments, ultimately resulting in the release of the genomic RNA. It is believed that native virions first lose, to various degrees, the innermost capsid protein VP4, resulting in the generation of subviral A-particles[93]; these are further converted to B-particles after the release of the RNA. Upon ultracentrifugation, native virions sediment at 150S and subviral

A-particles sediment at 135S. The latter are most likely an intermediate of the uncoating process that culminates in RNA expulsion and generation of empty subviral B-particles with a sedimentation constant of 80S. Subviral particles can also be produced *in vitro*. Exposure to pH 5.0[28,94] results primarily in 135S particles, and heating to 50–56°C for some minutes preferentially generates 80S particles (Lonberg-Holm, 1973).[94,97] This structural modification is accompanied by changes in antigenicity manifesting in the appearance of new (C-antigenic) epitopes that can be detected with specific antisera and monoclonal antibodies (Lonberg-Holm, 1973).[92,98] As explicitly shown for HRV2, A-particles are hydrophobic whereas B-particles are not.[7,8] This disagrees with the structural data on empty capsids that suggest the presence of the hydrophobic N-terminus of VP1 at the viral surface.[99–101]

Conversion of Major Group HRVs In Vitro

In vitro, interaction of soluble two-domain ICAM-1 (sICAM) with some major group viruses (e.g. HRV3 and HRV14, but not HRV16) above 20°C destabilizes the capsid to different extents in a time-dependent manner, giving rise to predominantly either 135S or 80S particles.[102–104] The tip of ICAM-1 is believed to first attach to the floor and south wall of the canyon with rather low affinity. Upon expulsion of the pocket factor, if present, more extensive interactions with both canyon walls become established; the latter requires physiologic temperature.[105,106] Concomitantly, the canyon undergoes conformational adjustments to better accommodate the receptor. Determination of the attachment of sICAM-1 to immobilized HRV3 and HRV14 by Biacore instrumentation demonstrated biphasic kinetics.[83,107] In principle, this supports the abovementioned mechanism; however, both kinetic components of the reaction were observed between 10°C and 25°C, temperatures where the structural changes should not occur. ICAM-1 presumably acts like an enzyme in diminishing the activation energy of the transition from the native virion to the subviral particle.[108] In addition to the receptor, low pH, which prevails in endosomes, may affect these structural alterations catalyzed by ICAM-1. When HRV16-sICAM-1 complexes preformed at 37°C were exposed to pH 6.0, the uncoating reaction was cooperative. For HRV3 that

possesses a pocket factor, a similar but only additive effect was observed. These data agree with the *in vivo* findings that low endosomal pH and the receptor act in concert in viral uncoating, similarly depending on the serotype.[104] Consequently, raising the endosomal pH to neutrality by lysosomotropic agents or V-ATPase inhibitors completely blocks infection of HRV16, but not of HRV3. For HRV14, the serotype with intermediate stability, the results are conflicting; this might be due to partial occupancy of the pocket.[104,109,110] Irrespective of the serotype, infection only occurs above 26°C.[111]

Conversion of Major Group HRVs In Vivo

Given the lack of lysosomal degradation of HRV14[85] when internalized at 34°C, the virus might use compartments of the recycling pathway for conformational change and uncoating. Nevertheless, at this temperature, essentially no virus was found in the endosomes by immunofluorescence microscopy and subcellular fractionation. However, at 20°C, where the virus remains in its native conformation, HRV14 was localized in early and late endosomes.[84] Thus, the structural alteration and uncoating may occur in these compartments. HRV3 and HRV14 can infect HeLa cells in the presence of bafilomycin.[104,109] Since this drug arrests cargo en route to lysosomes in early endosomes, at least these particular serotypes can undergo the receptor-catalyzed modification and uncoating in early endosomes (Fig. 5). As to which endosomal subcompartment the structural modification of the distinct major group serotypes will take place in, will depend on the time post infection and the pH required for this process. Since transit through the early endosomes is rapid (2–5 min) but residence in ECV/late endosomes amounts to 10–20 min,[70,73] the structural alteration most likely occurs in late compartments.

Conversion of Minor Group HRVs In Vitro

For HRV2, and presumably for the other minor group viruses, the *in vitro* conversion to subviral particles (mainly A-particles) exhibits a steep pH dependence that is unaffected by the presence of the cellular receptors. The kinetics of this modification can by analyzed using an

HRV2 conformation-specific monoclonal antibody (2G2) that detects C-antigenic but not native virions. While essentially no C-antigenic particles are observed at pH 5.8, 50% conversion is seen at pH 5.6, and all virions are converted to C-antigen below pH ≤ 5.4.[28,91] The conversion is virtually independent of the temperature within a range of 4°C to 34°C.

Conversion of Minor Group HRVs In Vivo

In vivo, formation of subviral particles and infection of HRV2 are solely low pH dependent as they are blocked when the endosomal pH is raised to neutrality and take place even when antibody-complexed virus is internalized via Fc-receptors.[91,112] Furthermore, the structural alteration is temperature-independent, provided that the virus is still routed into compartments that maintain a pH below the threshold (pH ≤ 5.6) for the conformational modification (ECV, late endosomes; above 16°C).[92] The products of the conversion, C-antigenic subviral particles, are partially recycled into the medium ("eluted particles").

To define the compartment where the conformational change of HRV2 occurs, we have taken advantage of various reagents known to affect distinct endocytic transport steps (see above).[70,84,92] In the presence of nocodazole, HRV2 accumulated in ECV. Nevertheless, the pH-dependent conformational change, recycling of C-antigenic particles, and infectivity were unaffected, but lysosomal degradation was prevented. Therefore, (1) ECV and late endosomes have a pH < 5.6 in HeLa cells as they induce the conformational modification of the virus; (2) recycling of C-antigenic ("eluted") viral particles can occur from ECV; and (3) uncoating takes place in ECV. This is in agreement with virus-receptor dissociation in mildly acidic early endosomes and the conformational modification occurring subsequently in more acidic EVC (Fig. 5).

Mechanism of Genome Uncoating: How is the Viral RNA Delivered into the Cytoplasm?

A crucial step in the virus entry pathway is membrane penetration/uncoating of the viral genome to allow for subsequent virus replication.

As non-enveloped viruses lack a membrane bilayer, they cannot fuse with a cellular membrane to deliver their genome or modified viral particles into the cytoplasm of the host cell. Thus, two potential mechanisms for membrane penetration of non-enveloped viruses have been proposed: (1) virus-induced membrane rupture and (2) selective genome translocation across a cellular membrane through pores formed by viral proteins (Fig. 6). The first mechanism (Fig. 6A) is well characterized for adenovirus, a DNA virus.[82] Adenovirus enters cells by receptor-mediated, clathrin-dependent endocytosis. The viral protease is activated within the reductive environment of endosomes. This results

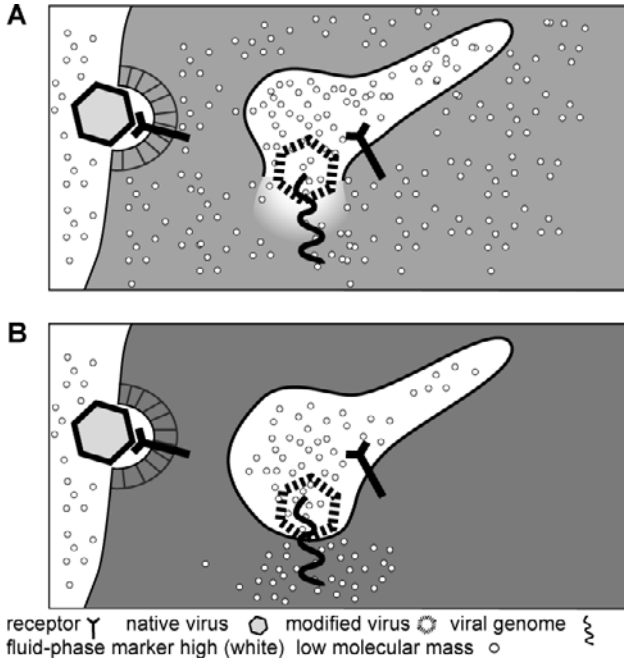


Fig. 6. Analysis of the mechanism of uncoating and penetration of non-enveloped viruses by endosome rupture or by pore formation. (**A**) Endosome rupture of adenovirus, for example, results in release of co-internalized fluid-phase marker into the cytosol. Fluid-phase markers of low and high molecular mass are lost from the endosomes to a similar extent. (**B**) Formation of virus-induced pores results in loss of low but not of high molecular mass markers from otherwise intact endosomes. It is assumed that the minor group HRV genome is transferred into the cytosol via such a size-restricting pore.

in the release of some viral proteins from the capsid and, with the aid of the co-receptor ($\alpha_v\beta_3$ or β_5-integrins), in rupture of the endosomal membrane. However, endosome lysis by adenovirus does not require low endosomal pH.[113] Upon endosome rupture, modified viral particles, together with the endosomal content, are delivered into the cytoplasm. The modified virions are then transported along microtubules to the nuclear pore complex, where uncoating takes place and the viral DNA is imported into the nucleus. The second mechanism (Fig. 6B) is assumed to account for uncoating and RNA translocation of poliovirus (Pv), another member of the picornavirus family, based primarily on *in vitro* structural and electrophysiological studies.[114–116] Similar to major group HRV, the Pv receptor (Pvr) catalyzes the structural alteration to 135S particles for infection. Upon interaction of Pv with soluble Pvr, the N-terminus of VP1 (a predicted amphipathic helix) is externalized. Thus, five amphipathic N-terminal helices together with five copies of VP4 might form the hypothetical channel in the target membrane through which the RNA might be released into the cytoplasm.[117] This view is supported by the recent cryo-EM image reconstruction of Pv attached to the membrane via five copies of Pvr showing a perturbation of the membrane just below the virus.[118] The importance of VP1 and of distinct amino acid residues in VP4 together with the N-terminal myristoyl-modification in uncoating was shown *in vivo* by using Pv mutants.[114,119,120] Pv with a mutation in VP4 that prevented RNA delivery into the cytoplasm and thus infection failed to form ion channels in artificial lipid bilayers.[114] However, the electrophysiological data do not differentiate whether *in vivo* Pv RNA is transported through such a channel across a cell membrane or whether the channel destabilizes the membrane and results in cytoplasmic delivery of the viral RNA together with the modified particles. In addition, it should be noted that the cellular site of Pv structural alteration and uncoating is still elusive.

Analysis of the Mechanism of Viral Genome Penetration into the Cytoplasm

We have recently established various assays to analyze and quantify the mechanism of viral genome penetration into the cytoplasm *in vivo* as well as *in vitro*.[113,121,122] The *in vivo* assay is based on the properties of

internalized fluid-phase markers that non-specifically label all endocytic vesicles, but preferentially endosomes on their way to lysosomes, depending on the internalization conditions such as time and temperature. Since fluid-phase markers (e.g. dextrans) do not bind to cellular membranes, they are released into the cytoplasm when endosomes are lysed or pores are formed that are large enough to allow passage of the markers. Internalized markers will be exposed to the low pH environment of intact endosomes, whereas in the presence of membrane-disrupting or pore-forming viruses, they will access the pH neutral cytoplasm (Fig. 6A). Using fluid-phase markers of different molecular mass, release of the marker by endosome lysis can be differentiated from release via a size-selective pore (Fig. 6A and B). When a pH-sensitive (FITC) and a pH-insensitive (Cy5) derivative of the same fluid-phase marker are internalized, flow cytometry of intact cells can be used for quantification of marker uptake (reflected by Cy5 fluorescence) and determination of the pH of the respective compartment (reflected by the ratio of FITC and Cy5 fluorescence). Detection of pH increase by this method indicates leakage of the markers from acidic endosomes into the pH-neutral cytosol. The number and pH of fluorescent endosomes can then be determined by subjecting cell homogenates to single organelle flow analysis (SOFA). Using fluid-phase markers of low and high molecular mass, a differentiation can be made between virus-induced endosome rupture and the formation of pores of a defined size appearing during genome penetration. When adenovirus was internalized for 20 min, this method revealed lysis of 40% of the endosomes that had been labeled either with low or high molecular mass markers.[113]

Penetration of Major Group HRVs

Penetration of modified particles of the major group virus HRV14 into the cytoplasm occurs by rupturing the endosomal membrane (Figs. 5 and 7). This is based on the following observations: HRV14 capsid proteins are detected in (isolated) endosomes under conditions that block their conformational modification (20°C) and thus uncoating. No virus was present in isolated endosomes at 34°C, but 135S and 80S particles were found in the cytosol.[84] Using our FACS and

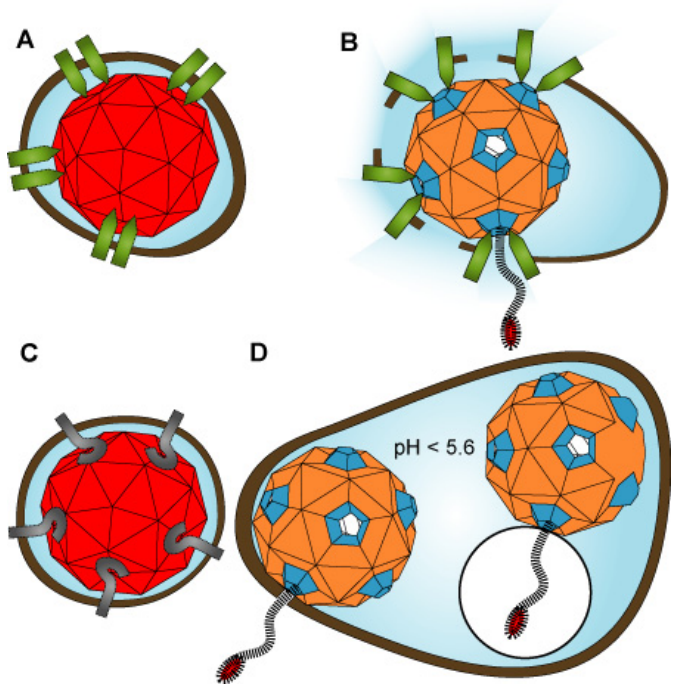


Fig. 7. Potential mechanism of uncoating and genome penetration of the major (**A, B**) and minor group (**C, D**) HRVs. (**A, B**): Binding of major group HRVs (e.g. HRV3, 14, 16) to ICAM-1 (shown in green) leads to receptor clustering and therefore internalization of the complex via clathrin-coated vesicles. The clathrin coat is rapidly lost (within 30 sec). The resulting vesicles (**A**) then fuse with early endosomes (**B**). The receptor-induced structural alteration results in externalization of hydrophobic residues. Concomitantly, the endosomal membrane is ruptured and the RNA is released into the cytoplasm (**B**). Given a residence time of 5 min in early and of 15–20 min in ECV/late endosomes, RNA penetration can occur from either compartment as a function of time and pH required for uncoating of the respective serotype. (**C, D**): Minor group viruses (e.g. HRV2) remain bound to LDLR (grey) in endocytic vesicles that maintain a pH > 6.0 (plasma membrane derived clathrin-coated and non-coated vesicles; (**C**)). After receptor dissociation at pH ≤ 6.0 (in early endosomes), the virus is transferred to endosomal carrier vesicles (ECV). The structural modification is solely induced by the low pH ≤ 5.6 in multivesicular ECV (**D**). The hydrophobic N-termini of VP1 together with VP4 might form a pore in the endosomal membrane through which the RNA is translocated into the cytoplasm. During this process, ECV remain intact and also maintain their low pH.

SOFA assays, the number of fluorescent endosomes was decreased by 20% in the presence of HRV14 with small and large markers being released equally from the endosomes (unpublished observations). Thus, HRV14 exhibits similar properties with respect to leakage of endosomal markers as the endosome-disrupting adenovirus.

Penetration of Minor Group HRVs

As pointed out above, the RNA is released from the minor group virus HRV2 by pore formation in ECV/late endosomes (Figs. 5 and 7). Our experimental data clearly support this mechanism in contrast to endosome rupture: (1) under conditions where complete uncoating has occurred, HRV2 capsid proteins can be localized in ECV/late endosomes[84]; (2) mainly C-antigenic 80S particles are found in isolated endosomes[92]; (3) FACS and SOFA assays demonstrated that HRV2 did not affect the integrity, number and pH of endosomes during uncoating and RNA penetration *in vivo*[113]; (4) selective passage of low-molecular mass dextran into the cytosol was indicative of the formation of a pore of limited size; and (5) using our *in vitro* "endosome leakage assay,"[121] specific pore formation induced by HRV2 upon endosome acidification was shown. Furthermore, peptides derived from the N-terminus of VP1 also induced such pores. These were similar in size as those that were caused by the GALA-INF3 peptide, an artificial acidic amphipathic peptide, where 16 amino acids were replaced by the N-terminal residues of the fusion peptide of influenza HA2. These data demonstrate the importance of the N-terminus of VP1 in RNA translocation into the cytosol and further suggest that the derived 24-amino acid long oligopeptide might form channels of 5 to 10 Å diameter in the endosomal membrane as calculated for similar viral peptides.[121,123] Taken together, this once more documents the differences in the mode of entry, conformational change, uncoating and RNA penetration of major and minor group HRV.

HRV Uncoating at the Plasma Membrane

At least for HRV2, we have clearly shown that infection can occur from the plasma membrane when it is exposed to low pH.[91] HRV2, attached

to its receptors on the HeLa cell surface and exposed to a pH 5.3 buffer at 4°C, is entirely converted to C-antigen. These C-antigenic particles are then gradually released from the plasma membrane. When cells were subsequently warmed to 34°C after dissipation of the low endosomal pH by bafilomycin to inhibit any endosomal uncoating but to permit virus replication, viral *de novo* synthesis was observed.[91] This demonstrates that HRV2 uncoating and RNA transfer into the cytosol can be artificially induced at the plasma membrane by exposure to low pH. Apparently, the major group virus HRV14 fails to uncoat at the plasma membrane at neutral pH given the lack of infection of HeLa cells over-expressing non-functional mutant dynamin.[81] This is in agreement with RNA penetration into the cytosol by rupture of a cellular (endosomal) membrane, a process deleterious for the host cell when it occurs at the plasma membrane.

Potential Mechanisms of RNA Egress

In cryo-EM pictures taken from HRV3-sICAM-1 complexes formed at 4°C followed by heating to 37°C, the receptor is seen to remain bound even after RNA egress.[101] The structural data derived from the image analysis were interpreted as indicating that the breathing capsid was held in an "open conformation" by the receptor acting like a wedge. This is particularly visible at the star-shaped dome at the five-fold axes that become enlarged and raise upward, resulting in an overall increase in diameter of the particle by 4% and a loosening of the inter-protomer contacts; the density at the pseudo-threefold axes was interpreted as indicating the externalization of VP4 and of the N-terminus of VP1 close to the ICAM-1 molecule, which would even provide a hydrophobic residue for interaction.

The cryo-EM structure of empty HRV14 particles generated in the absence of a receptor by heating to 55°C showed similar rearrangements. However, the density was not increased at the pseudo three-fold axes but rather at the five-fold axes, thereby suggesting exit of the VP1 N-terminus together with VP4 at this location.[99] The presence of VP4 is in agreement with biochemical data indicating that a large percentage of VP4 remained associated with subviral particles.[93] In both

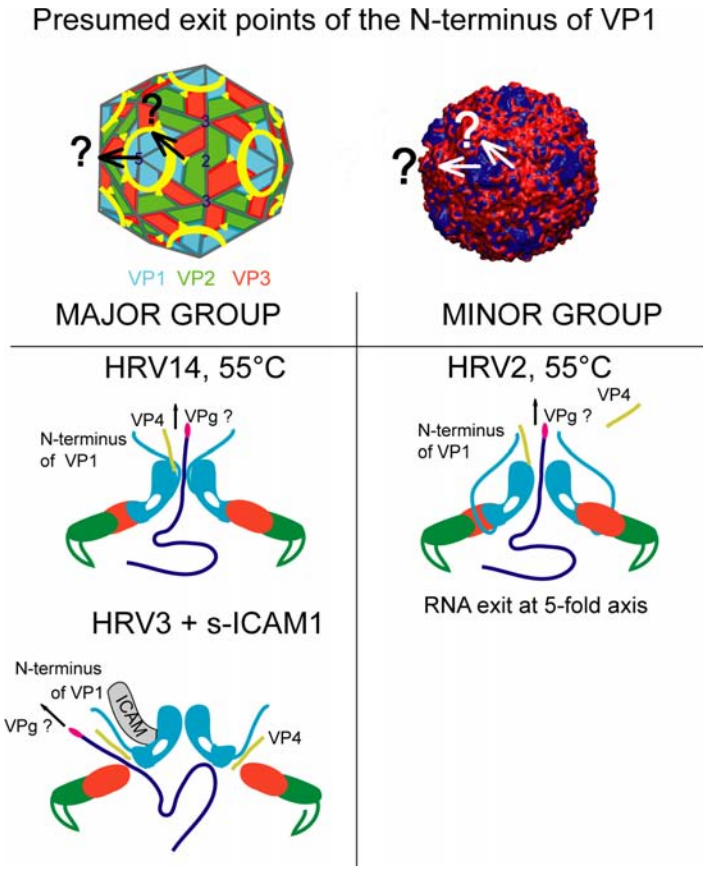


Fig. 8. Externalization of the N-terminus of VP1 and exit of the RNA. A schematic drawing of the virus (*left*) illustrates the positions of 2, 3, 5, and pseudo three-fold axes (yellow triangle) and of the viral canyon; VP1, blue; VP2, green; VP3; red. Density maps of native HRV2 were calculated from the X-ray data (red) and superimposed onto the cryo-EM reconstructions of empty HRV2 particles (blue). Regions in blue in the superposition indicate where the empty particle is enlarged with respect to the native virus. This is particularly pronounced at the five-fold axes (*right*). Possible points of externalization of the N-terminus of VP1 are indicated. Cryo-EM of empty HRV14 capsids suggests exit at the five-fold axis, whereas in a complex between empty HRV3 and sICAM–1 VP1 appears to externalize at the pseudo three-fold axis. Neither of the empty major group HRV shows enlargement of the VP3 β-cylinder as seen in HRV2. RNA exit might thus take place through any of the enlarged β-cylinders in HRV2, but in HRV3 and HRV14 only one of the

reconstructions, no expansion of the VP3 β-cylinder was seen. Such an expansion would be required to generate a channel sufficiently large for RNA egress. It is thus believed that the β-cylinder of at least one, but not of all 12, five-fold axes enlarges to allow for transit of the RNA (Fig. 8). Since cryo-EM image reconstruction is an averaging technique, changes at single vertices would escape detection. The data might indicate that, *in vivo*, the A-particle is handed over from ICAM-1 to the membrane, where it directly interacts with the lipids via the hydrophobic N-terminus of VP1, possibly aided by the myristoylated N-terminus of VP4. However, it should be kept in mind that the rigid ICAM-1 molecule has a length of ~19 nm and, consequently, the virus would be held at this distance from the endosomal membrane, thus making such a scenario unlikely. Rather, the movement of ICAM-1 by 30° concomitant with the conformational modification[101] might result in membrane destabilization. This process could be directly coupled with RNA uncoating and penetration into the cytosol (Fig. 7).

In the case of HRV2, cryo-EM reconstruction of empty capsids generated by heating to 55°C revealed, in addition to the 4% expansion of the capsid, the presence of density below the pseudo three-fold axis that was interpreted as stemming from the exiting N-terminus of VP1. However, there was no density attributable to VP4, suggesting that it was completely absent.[100] However, in HRV2, the VP3 β-cylinders had expanded, making it highly probable that the RNA can exit at any of the 12 five-fold axes (Fig. 8). The high flexibility of LDLR and LRP and the presumed attachment to the ligand-binding domain via several ligand-binding modules wound around each five-fold axis might, in principle, permit the virus to come close to the membrane. This would permit contacts between the N-termini of VP1 and the endosomal membrane to be established with the receptor still remaining attached to the virus.

---

12 β-cylinders might enlarge and give way for the RNA. VP4 is probably almost entirely lost in HRV2, but VP4 might remain bound to HRV14 at the five-fold axes or at the pseudo three-fold axes (HRV3). Schemes are based on data from Hewat (empty HRV2[100] and empty HRV14[99]) and from Xing (empty HRV3 with bound soluble two-domain ICAM-1[101]). RNA release might start with the 5′-terminus as indicated by Vpg, but recent studies with atomic force microscopy suggest instead that the 3′-end exits first.[124]

However, plasma membrane-bound HRV2 was found to dissociate at pH 6.0, whereas transformation into A-particles requires pH ≤ 5.6.[91] Therefore, the structural change occurs after the virus is dissociated from the receptor and intimate membrane contacts can take place in ECV, presumably by insertion of the VP1 N-termini into the lipid bilayer (Fig. 7). Since ECV have many internal vesicles, viral proteins may insert into such internal membranes, resulting in RNA translocation into their lumen. These internal membranes are then degraded together with the viral RNA in lysosomes.[85] At least in the case of HRV2, this would lead to unproductive uncoating and may explain the low ratio of infectious particles to total particles (1:24 up to 1:240).[9]

What triggers the extrusion of the RNA? Recent investigations by atomic force microscopy (AFM) of RNA cores have shown, for the related poliovirus, that the RNA remains in a compact conformation in the form of a sphere upon removal of the proteins.[125] This spherical structure unfolds upon heating, but it reforms as soon as the temperature is reduced. This can be taken to indicate that the compaction is encoded in the sequence, resulting in a multitude of hairpins with different lengths. It is conceivable that upon encapsidation, one of the termini of the RNA is left at a preferred location, presumably close to one of the five-fold axes, from where it exits as soon as the conformational transition of the capsid takes place. AFM of HRV2 incubated at low pH revealed that RNA molecules seen in the process of exiting the virion lacked the short forks at their termini seen in RNAs completely separated from the virus particles. These structures presumably represent the cloverleaf structure at 5′ end; if this is correct, RNA exit starts with the 3′ end and not, as indicated in Fig. 8, with the small genome-linked protein, VPg, that is covalently attached to the 5′-end of the RNA.[124] Assuming the equivalence of all 12 vertices of the icosahedron, RNA expulsion could occur at any of these. It might be hypothesized that, for HRV2, the site of exit must be in close apposition to the endosomal membrane for successful transfer into the cytosol. Therefore, the probability of a suitable arrangement at the membrane might be not much more than 1:12. This, in addition, may account for the low infectious particle to total particle ratio of rhinoviruses.

What is Known of Receptor Expression and HRV Infection of the Airway Epithelium?

The upper airway epithelium (nasal cavities and nasopharynx) is the major site of HRV replication,[126] but occasionally the lower airways (tracheobronchial tree) also become infected.[127–130] In any case, no marked cytopathic effect is observed.

The respiratory epithelium is built from different cell types,[131] predominantly from ciliated columnar (epithelial) cells (the most abundant) and mucous goblet cells. Both are polarized with an apical and a basolateral plasma membrane separated by tight junctions. The basal (short) cells are small and rounded and are in contact with the basal lamina; they can differentiate into the other cell types.

In bronchial tissue infected *ex vivo* with the major group HRV16, only a small subset of the epithelial cells shows virus replication.[128] Whether this is related to receptor expression is not clear. The few data that are available on the expression of ICAM-1 in normal nasal epithelium *in situ* are contradictory. Whereas in earlier studies ICAM-1 was not found, more recent work demonstrated the presence of the receptor in both ciliated and basal cells in inferior turbinates.[132] In nasal polyps and nasopharyngeal lymph nodes, receptor expression is predominantly restricted to the basal cells and only traces are found at the ciliated surface of columnar cells. Thus, with increasing differentiation of the basal cells to columnar ciliated cells, expression of ICAM-1 decreases.[133,134] Using primary cultures of polarized tracheal epithelial cells, ICAM-1 was found on both the apical and basolateral plasma membrane, although total expression was low as compared to HeLa cells.[128,135] Furthermore, when such cells are cultured under conditions that permit full differentiation into ciliated cells, ICAM-1 expression decreases.[136] However, major group HRVs preferentially replicate in less differentiated tracheal epithelial cells. These studies suggest that susceptibility to infection by major group HRVs correlates with the level of ICAM-1 expression. Nevertheless, expression of ICAM-1 in the nasal epithelium has not been quantified *in situ* and the polarity of expression is unknown.

No data have been published on the expression and localization of receptors for minor group HRVs in the airway epithelium *in situ*. LDLR expression has been solely studied in human tracheal epithelial cell cultures. Similar to major group HRVs that stimulate ICAM-1 expression, HRV2 infection resulted in up-regulation of LDLR mRNA and protein.[137,138] HRV2 infection was completely blocked by an antibody against LDLR in tracheal cells. However, the situation might differ in nasal epithelia, where VLDLR and/or LRP might be present and implicated in viral attachment. In any case, the polarity of receptor expression and virus entry and release are unexplored. It is of note that infections by minor group HRV occur more frequently than expected from the relative numbers of serotypes belonging to the two groups. This might be taken to indicate higher concentrations and/or higher efficiency of these receptors with respect to binding and internalization as compared to ICAM-1.

Concluding Remarks

Despite great effort to elucidate the early events of infection, such as attachment to the host cells, penetration, and release of the viral genome into the cytosol, many questions have remained open: (1) the basis for the strict receptor specificity is unclear; (2) the structure, localization and mode of receptor binding of major and minor group HRV *in vivo* is completely unknown; (3) it is not understood why and how HRVs penetrate into the cytoplasm by using different pathways; (4) it is poorly understood why major group serotypes differ in their requirement for low pH for structural modification and uncoating; (5) it is not known in detail how the RNA genome is released and transferred into the cytosol; and, most importantly, (6) the mechanism of infection of the airway epithelium is largely unexplored with respect to receptor expression *per se*, polarity of virus entry and vectorial release of infectious virions.

Acknowledgments

This work was supported by grants from the Virology Foundation (D.B. and R.F.) and from the Austrian Science Foundation (P-12967

and P-17590 to R.F. and P-16699-B10 to D.B.). We particularly thank I. Fita, E. Hewat, T. Hovi, E. Kenndler, and N. Verdaguer for their collaboration and Simone Fuchs for her help with the artwork.

References

1. Tyrell AD, Fielder M. (2003) *Cold wars. The fight against the common cold*. Oxford University Press.
2. Lewis JK, Bothner B, Smith TJ, Siuzdak G. (1998) Antiviral agent blocks breathing of the common cold virus. *Proc Natl Acad Sci USA* **95**: 6774–6778.
3. Phelps DK, Post CB. (1999) Molecular dynamics investigation of the effect of an antiviral compound on human rhinovirus. *Protein Sci* **8**: 2281–2289.
4. Phelps DK, Rossky PJ, Post CB. (1998) Influence of an antiviral compound on the temperature dependence of viral protein flexibility and packing: a molecular dynamics study. *J Mol Biol* **276**: 331–337.
5. Reisdorph N, Thomas JJ, Katpally U, *et al.* (2003) Human rhinovirus capsid dynamics is controlled by canyon flexibility. *Virology* **314**: 34–44.
6. Speelman B, Brooks BR, Post CB. (2001) Molecular dynamics simulations of human rhinovirus and an antiviral compound. *Biophys J* **80**: 121–129.
7. Lonberg-Holm K, Crowell RL, Philipson L. (1976) Unrelated animal viruses share receptors. *Nature* **259**: 679–681.
8. Lonberg-Holm K, Gosser LB, Shimshick J. (1976) Interaction of liposomes with subviral particles of poliovirus type 2 and rhinovirus type 2. *J Virol* **19**: 746–749.
9. Abraham G, Colonno RJ. (1984) Many rhinovirus serotypes share the same cellular receptor. *J Virol* **51**: 340–345.
10. Colonno RJ, Callahan PL, Long WJ. (1986) Isolation of a monoclonal antibody that blocks attachment of the major group of human rhinoviruses. *J Virol* **57**: 7–12.
11. Tomassini JE, Colonno RJ. (1986) Isolation of a receptor protein involved in attachment of human rhinoviruses. *J Virol* **58**: 290–295.
12. Greve JM, Davis G, Meyer AM, *et al.* (1989) The major human rhinovirus receptor is ICAM-1. *Cell* **56**: 839–847.
13. Staunton DE, Merluzzi VJ, Rothlein R, *et al.* (1989) A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. *Cell* **56**: 849–853.

14. Tomassini E, Graham T, DeWitt C, *et al.* (1989) cDNA cloning reveals that the major group rhinovirus receptor on HeLa cells is intercellular adhesion molecule 1. *Proc Natl Acad Sci USA* **86**: 4907–4911.
15. Gruenberger M, Wandl R, Nimpf J, *et al.* (1995) Avian homologs of the mammalian low-density lipoprotein receptor family bind minor receptor group human rhinovirus. *J Virol* **69**: 7244–7247.
16. Hofer F, Gruenberger M, Kowalski H, *et al.* (1994) Members of the low density lipoprotein receptor family mediate cell entry of a minor-group common cold virus. *Proc Nat Acad Sci USA* **91**: 1839–1842.
17. Marlovits TC, Abrahamsberg C, Blaas D. (1998) Soluble LDL minireceptors. Minimal structure requirements for recognition of minor group human rhinovirus. *J Biol Chem* **273**: 33835–33840.
18. Marlovits TC, Abrahamsberg C, Blaas D. (1998) Very-low-density lipoprotein receptor fragment shed from HeLa cells inhibits human rhinovirus infection. *J Virol* **72**: 10246–10250.
19. Blomqvist S, Savolainen C, Raman L, *et al.* (2002) Human rhinovirus 87 and enterovirus 68 represent a unique serotype with rhinovirus and enterovirus features. *J Clin Microbiol* **40**: 4218–4223.
20. Laine P, Savolainen C, Blomqvist S, Hovi T. (2005) Phylogenetic analysis of human rhinovirus capsid protein VP1 and 2A protease coding sequences confirms shared genus-like relationships with human enteroviruses. *J Gen Virol* **86**: 697–706.
21. Savolainen C, Laine P, Mulders MN, Hovi T. (2004) Sequence analysis of human rhinoviruses in the RNA-dependent RNA polymerase coding region reveals large within-species variation. *J Gen Virol* **85**: 2271–2277.
22. Vlasak M, Roivainen M, Reithmayer M, *et al.* (2005) The minor receptor group of human rhinovirus (HRV) includes HRV23 and HRV25, but the presence of a lysine in the VP1 HI loop is not sufficient for receptor binding. *J Virol* **79**: 7389–7395.
23. Ledford RM, Patel NR, Demenczuk TM, *et al.* (2004) VP1 sequencing of all human rhinovirus serotypes: insights into genus phylogeny and susceptibility to antiviral capsid-binding compounds. *J Virol* **78**: 3663–3674.
24. Mckinlay MA, Pevear DC, Rossmann MG. (1992) Treatment of the picornavirus common cold by inhibitors of viral uncoating and attachment. *Ann Rev Microbiol* **46**: 635–654.
25. Kim KH, Willingmann P, Gong ZX, *et al.* (1993) A comparison of the anti-rhinoviral drug binding pocket in HRV14 and HRV1A. *J Mol Biol* **230**: 206–227.

26. Andries K, Dewindt B, Snoeks J, *et al.* (1990) Two groups of rhinoviruses revealed by a panel of antiviral compounds present sequence divergence and differential pathogenicity. *J Virol* **64**: 1117–1123.
27. Verdaguer N, Blaas D, Fita I. (2000) Structure of human rhinovirus serotype 2 (HRV2). *J Mol Biol* **300**: 1179–1194.
28. Gruenberger M, Pevear D, Diana GD, *et al.* (1991) Stabilization of human rhinovirus serotype-2 against pH-induced conformational change by antiviral compounds. *J Gen Virol* **72**: 431–433.
29. Okun VM, Nizet S, Blaas D, Kenndler E. (2002) Kinetics of thermal denaturation of human rhinoviruses in the presence of anti-viral capsid binders analyzed by capillary electrophoresis. *Electrophoresis* **23**: 896–902.
30. Zhang AQ, Nanni RG, Li T, *et al.* (1993) Structure determination of antiviral compound SCH-38057 complexed with human rhinovirus-14. *J Mol Biol* **230**: 857–867.
31. Rossmann MG. (1994) Viral cell recognition and entry. *Protein Sci* **3**: 1712–1725.
32. Savolainen C, Blomqvist S, Mulders MN, Hovi T. (2002) Genetic clustering of all 102 human rhinovirus prototype strains: serotype 87 is close to human enterovirus 70. *J Gen Virol* **83**: 333–340.
33. Sherry B, Mosser AG, Colonno J, Rueckert RR. (1985) Use of monoclonal antibodies to identify four neutralization immunogens on a common cold picornavirus human rhinovirus 14. *J Virol* **57**: 246–257.
34. Sherry B, Rueckert R. (1985) Evidence for at least two dominant neutralization antigens on human rhinovirus 14. *J Virol* **53**: 137–143.
35. Rossmann MG. (1989) The canyon hypothesis. Hiding the host cell receptor attachment site on a viral surface from immune surveillance. *J Biol Chem* **264**: 14587–14590.
36. Olson NH, Kolatkar PR, Oliveira MA, *et al.* (1993) Structure of a human rhinovirus complexed with its receptor molecule. *Proc Natl Acad Sci USA* **90**: 507–511.
37. Chapman MS, Rossmann MG. (1993) Comparison of surface properties of picornaviruses: strategies for hiding the receptor site from immune surveillance. *Virology* **195**: 745–756.
37a. Laine P, Blomqvist S, Savolainen C, *et al.* (2006) Alignment of capsid protein VP1 sequnces of all human rhinovirus prototype strains: conserved motifs and functional domains. *J Gen Virol* **8**: 129–138.
38. Smith TJ, Chase ES, Schmidt TJ, *et al.* (1996) Neutralizing antibody to human rhinovirus 14 penetrates the receptor-binding canyon. *Nature* **383**: 350–354.

39. Kim S, Smith TJ, Chapman MS, *et al.* (1989) Crystal structure of human rhinovirus serotype-1A (HRV1A). *J Mol Biol* **210**: 91–111.
40. Oliveira MA, Zhao R, Lee WM, *et al.* (1993) The structure of human rhinovirus 16. *Structure* **1**: 51–68.
41. Rossmann M, Arnold E, Erickson J, *et al.* (1985) Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* **317**: 145–153.
42. Zhao R, Pevear DC, Kremer MJ, *et al.* (1996) Human rhinovirus 3 at 3.0 angstrom resolution. *Structure* **4**: 1205–1220.
43. Bella J, Kolatkar PR, Marlor CW, *et al.* (1998) The structure of the two amino-terminal domains of human ICAM-1 suggests how it functions as a rhinovirus receptor and as an LFA-1 integrin ligand. *Proc Natl Acad Sci USA* **95**: 4140–4145.
44. Rudenko G, Deisenhofer J. (2003) The low-density lipoprotein receptor: ligands, debates and lore. *Curr Opin Struct Biol* **13**: 683–689.
45. Staunton DE, Dustin ML, Erickson HP, Springer TA. (1990) The arrangement of the immunoglobulin-like domains of ICAM-1 and the binding sites for LFA-1 and rhinovirus. *Cell* **61**: 243–254.
46. Rothlein R, Dustin ML, Marlin SD, Springer TA. (1986) A human intercellular adhesion molecule (ICAM-1) distinct from LFA-1. *J Immunol* **137**: 1270–1274.
47. Casasnovas JM, Bickford JK, Springer TA. (1998) The domain structure of ICAM-1 and the kinetics of binding to rhinovirus. *J Virol* **72**: 6244–6246.
48. Fass D, Blacklow S, Kim PS, Berger JM. (1997) Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module. *Nature* **388**: 691–693.
49. Rudenko G, Henry L, Henderson K, *et al.* (2002) Structure of the LDL receptor extracellular domain at endosomal pH. *Science* **298**: 2353–2358.
50. Russell DW, Schneider WJ, Yamamoto T, *et al.* (1984) Domain map of the LDL receptor: sequence homology with the epidermal growth factor precursor. *Cell* **37**: 577–585.
51. Strickland DK, Kounnas MZ, Williams SE, Argraves WS. (1994) LDL receptor-related protein (LRP): a multiligand receptor. *Fibrinolysis* **8**: 204–215.
52. Bhella D, Goodfellow IG, Roversi P, *et al.* (2004) The structure of echovirus type 12 bound to a two-domain fragment of its cellular

attachment protein decay-accelerating factor (CD 55). *J Biol Chem* **279**: 8325–8332.

53. He Y, Bowman VD, Mueller S, *et al.* (2000) Interaction of the poliovirus receptor with poliovirus. *Proc Natl Acad Sci USA* **97**: 79–84.
54. He Y, Chipman PR, Howitt J, *et al.* (2001) Interaction of coxsackievirus B3 with the full length coxsackievirus-adenovirus receptor. *Nat Struct Biol* **8**: 874–878.
55. Neumann E, Moser R, Snyers L, *et al.* (2003) A cellular receptor of human rhinovirus type 2, the very-low-density lipoprotein receptor, binds to two neighboring proteins of the viral capsid. *J Virol* **77**: 8504–8511.
56. Xing L, Huhtala M, Pietiainen V, *et al.* (2004) Structural and functional analysis of integrin alpha2I domain interaction with echovirus 1. *J Biol Chem* **279**: 11632–11638.
57. Xiao C, Tuthill TJ, Bator Kelly CM, *et al.* (2004) Discrimination among rhinovirus serotypes for a variant ICAM-1 receptor molecule. *J Virol* **78**: 10034–10044.
58. Verdaguer N, Fita I, Reithmayer M, *et al.* (2004) X-ray structure of a minor group human rhinovirus bound to a fragment of its cellular receptor protein. *Nat Struct Mol Biol* **11**: 429–434.
59. Xiao C, Bator CM, Bowman VD, *et al.* (2001) Interaction of coxsackievirus A21 with its cellular receptor, ICAM-1. *J Virol* **75**: 2444–2451.
60. Moser R, Wruss J, Snyers L, *et al.* (2005) Concatenation of VDL-receptor ligand-binding repeat 3 results in highly efficient inhibitors of HRV infection: evidence for multi-module binding. Submitted.

60a. Wruss J, Runzler D, Steiger C, *et al.* (2007) Attachment of VLDL receptors to an icosahedral virus along the 5-fold symmetry axis: multiple binding modes evidenced by fluorescence correlation spectroscopy. *Biochemistry* **46**: 6331–6339.

61. Conner SD, Schmid SL. (2003) Regulated portals of entry into the cell. *Nature* **422**: 37–44.
62. Pelkmans L, Helenius A. (2003) Insider information: what viruses tell us about endocytosis. *Curr Opin Cell Biol* **15**: 414–422.
63. Damm EM, Pelkmans L, Kartenbeck J, *et al.* (2005) Clathrin- and caveolin-1-independent endocytosis: entry of simian virus 40 into cells devoid of caveolae. *J Cell Biol* **168**: 477–488.
64. Hansen SH, Sandvig K, van Deurs B. (1993) Molecules internalized by clathrin-independent endocytosis are delivered to endosomes containing transferrin receptors. *J Cell Biol* **123**: 89–97.

65. Pelkmans L, Burli T, Zerial M, Helenius A. (2004) Caveolin-stabilized membrane domains as multifunctional transport and sorting devices in endocytic membrane traffic. *Cell* **118**: 767–780.
66. Mukherjee S, Ghosh RN, Maxfield FR. (1997) Endocytosis. *Physiol Rev* **77**: 759–803.
67. Nishi T, Forgac M. (2002) The vacuolar (H^+)-ATPases — nature's most versatile proton pumps. *Nat Rev Mol Cell Biol* **3**: 94–103.
68. Mellman I, Fuchs R, Helenius A. (1986) Acidification of the endocytic and exocytic pathways. *Ann Rev Biochem* **55**: 663–700.
69. Fuchs R, Ellinger A, Pavelka M, *et al.* (1994) Rat liver endocytic coated vesicles do not exhibit ATP-dependent acidification *in vitro*. *Proc Nat Acad Sci USA* **91**: 4811–4815.
70. Bayer N, Schober D, Prchla E, *et al.* (1998) Effect of bafilomycin A1 and nocodazole on endocytic transport in HeLa cells: implications for viral uncoating and infection. *J Virol* **72**: 9645–9655.
71. Clague MJ. (1998) Molecular aspects of the endocytic pathway. *Biochem J* **336**: 271–282.
72. Raiborg C, Rusten TE, Stenmark H. (2003) Protein sorting into multivesicular endosomes. *Curr Opin Cell Biol* **15**: 446–455.
73. Baravalle G, Schober D, Huber M, *et al.* (2005) Transferrin recycling and dextran transport to lysosomes is differentially affected by bafilomycin, nocodazole, and low temperature. *Cell Tissue Res* **320**: 99–113.
74. Sonawane ND, Verkman AS. (2003) Determinants of [Cl-] in recycling and late endosomes and Golgi complex measured using fluorescent ligands. *J Cell Biol* **160**: 1129–1138.
75. Aniento F, Emans N, Griffiths G, Gruenberg J. (1993) Cytoplasmic dynein-dependent vesicular transport from early to late endosomes. *J Cell Biol* **123**: 1373–1387.
76. Griffiths G, Hoflack B, Simons K, *et al.* (1988) The mannose 6-phosphate receptor and the biogenesis of lysosomes. *Cell* **52**: 329–341.
77. Maxfield FR, McGraw TE. (2004) Endocytic recycling. *Nat Rev Mol Cell Biol* **5**: 121–132.
78. Ghosh RN, Gelman DL, Maxfield FR. (1994) Quantification of low density lipoprotein and transferrin endocytic sorting HEp2 cells using confocal microscopy. *J Cell Sci* **107**: 2177–2189.
79. Sheff DR, Daro EA, Hull M, Mellman I. (1999) The receptor recycling pathway contains two distinct populations of early endosomes with different sorting functions. *J Cell Biol* **145**: 123–139.

80. Staunton DE, Gaur A, Chan PY, Springer TA. (1992) Internalization of a major group human rhinovirus does not require cytoplasmic or transmembrane domains of ICAM-1. *J Immunol* **148**: 3271–3274.
81. DeTulleo L, Kirchhausen T. (1998) The clathrin endocytic pathway in viral infection. *EMBO J* **17**: 4585–4593.
82. Meier O, Greber UF. (2003) Adenovirus endocytosis. *J Gene Med* **5**: 451–462.
83. Casasnovas M, Springer TA. (1995) Kinetics and thermodynamics of virus binding to receptor. Studies with rhinovirus, intercellular adhesion molecule-1 (ICAM-1), and surface plasmon resonance. *J Biol Chem* **270**: 13216–13224.
84. Schober D, Kronenberger P, Prchla E, *et al.* (1998) Major and minor receptor group human rhinoviruses penetrate from endosomes by different mechanisms. *J Virol* **72**: 1354–1364.
85. Lonberg-Holm K, Korant BD. (1972) Early interaction of rhinoviruses with host cells. *J Virol* **9**: 29–40.
86. Strickland DK, Gonias SL, Argraves WS. (2002) Diverse roles for the LDL receptor family. *Trends Endocrinol Metab* **13**: 66–74.
87. Bayer N, Schober D, Hüttinger M, *et al.* (2001) Inhibition of clathrin-dependent endocytosis has multiple effects on HRV2 cell entry. *J Biol Chem* **276**: 3952–3962.
88. Huber M, Brabec M, Bayer N, *et al.* (2001) Elevated endosomal pH in HeLa cells overexpressing mutant dynamin can affect infection by pH-sensitive viruses. *Traffic* **2**: 727–736.
89. Madshus IH, Sandvig K, Olsnes S, van Deurs B. (1987) Effect of reduced endocytosis induced by hypotonic shock and potassium depletion on the infection of Hep 2 cells by picornaviruses. *J Cell Physiol* **131**: 14–22.
90. Snyers L, Zwickl H, Blaas D. (2003) Human rhinovirus type 2 is internalized by clathrin-mediated endocytosis. *J Virol* **77**: 5360–5369.
91. Brabec M, Baravalle G, Blaas D, Fuchs R. (2003) Conformational changes, plasma membrane penetration, and infection by human rhinovirus type 2: role of receptors and low pH. *J Virol* **77**: 5370–5377.
92. Prchla E, Kuechler E, Blaas D, Fuchs R. (1994) Uncoating of human rhinovirus serotype 2 from late endosomes. *J Virol* **68**: 3713–3723.
93. Casasnovas JM, Springer TA. (1994) Pathway of rhinovirus disruption by soluble intercellular adhesion molecule 1 (ICAM-1): an intermediate in which ICAM-1 is bound and RNA is released. *J Virol* **68**: 5882–5889.

94. Korant BD, Lonberg-Holm K, Noble J, Stasny JT. (1972) Naturally occurring and artificially produced components of three rhinoviruses. *Virology* **48**: 71–86.
95. Lonberg-Holm K, Noble-Harvey J. (1973) Comparison of *in vitro* and cell-mediated alteration of a human rhinovirus and its inhibition by sodium dodecyl sulfate. *J Virol* **12**: 819–824.
96. Lonberg-Holm K, Yin FH. (1973) Antigenic determinants of infective and inactivated human rhinovirus type 2. *J Virol* **12**: 114–123.
97. Noble JN, Lonberg-Holm K. (1973) Interactions of components of human rhinovirus type 2 with HeLa cells. *Virology* **51**: 270–278.
98. Neubauer C, Frasel L, Kuechler E, Blaas D. (1987) Mechanism of entry of human rhinovirus 2 into HeLa cells. *Virology* **158**: 255–258.
99. Hewat EA, Blaas D. (2004) Cryoelectron microscopy analysis of the structural changes associated with human rhinovirus type 14 uncoating. *J Virol* **78**: 2935–2942.
100. Hewat EA, Neumann E, Blaas D. (2002) The concerted conformational changes during human rhinovirus 2 uncoating. *Mol Cell* **10**: 317–326.
101. Xing L, Casasnovas JM, Cheng RH. (2003) Structural analysis of human rhinovirus complexed with ICAM-1 reveals the dynamics of receptor-mediated virus uncoating. *J Virol* **77**: 6101–6107.
102. Greve J, Forte C, Marlor C, *et al.* (1991) Mechanisms of receptor-mediated rhinovirus neutralization defined by two soluble forms of ICAM-1. *J Virol* **65**: 6015–6023.
103. Hoover-Litty H, Greve JM. (1993) Formation of rhinovirus-soluble ICAM-1 complexes and conformational changes in the virion. *J Virol* **67**: 390–397.
104. Nurani G, Lindqvist B, Casasnovas JM. (2003) Receptor priming of major group human rhinoviruses for uncoating and entry at mild low-pH environments. *J Virol* **77**: 11985–11991.
105. Casasnovas JM. (2000) The dynamics of receptor recognition by human rhinoviruses. *Trends Microbiol* **8**: 251–254.
106. Rossmann MG, Bella J, Kolatkar PR, *et al.* (2000) Cell recognition and entry by rhino- and enteroviruses. *Virology* **269**: 239–247.
107. Xing L, Tjarnlund K, Lindqvist B, *et al.* (2000) Distinct cellular receptor interactions in poliovirus and rhinoviruses. *EMBO J* **19**: 1207–1216.
108. Tsang SK, McDermott BM, Racaniello VR, Hogle JM. (2001) Kinetic analysis of the effect of poliovirus receptor on viral uncoating: the receptor as a catalyst. *J Virol* **75**: 4984–4989.

109. Bayer N, Prchla E, Schwab M, *et al.* (1999) Human rhinovirus HRV14 uncoats from early endosomes in the presence of bafilomycin. *FEBS Lett* **463**: 175–178.
110. Perez L, Carrasco L. (1993) Entry of poliovirus into cells does not require a low-pH step. *J Virol* **67**: 4543–4548.
111. Lee WM, Monroe SS, Rueckert RR. (1993) Role of maturation cleavage in infectivity of picornaviruses: activation of an infectosome. *J Virol* **67**: 2110–2122.
112. Baravalle G, Snyers L, Brabec M, *et al.* (2004) Human rhinovirus type 2-antibody complexes enter and infect cells via Fc-gamma receptor IIB1. *J Virol* **78**: 2729–2735.
113. Brabec M, Schober D, Wagner E, *et al.* (2005) Opening of size-selective pores in endosomes during human rhinovirus serotype 2 *in vivo* uncoating monitored by single-organelle flow analysis. *J Virol* **79**: 1008–1016.
114. Danthi P, Tosteson M, Li QH, Chow M. (2003) Genome delivery and ion channel properties are altered in VP4 mutants of poliovirus. *J Virol* **77**: 5266–5274.
115. Hogle JM. (2002) Poliovirus cell entry: common structural themes in viral cell entry pathways. *Annu Rev Microbiol* **56**: 677–702.
116. Tosteson MT, Chow M. (1997) Characterization of the ion channels formed by poliovirus in planar lipid membranes. *J Virol* **71**: 507–511.
117. Belnap DM, Filman DJ, Trus BL, *et al.* (2000) Molecular tectonic model of virus structural transitions: the putative cell entry states of poliovirus. *J Virol* **74**: 1342–1354.
118. Bubeck D, Filman DJ, Hogle JM. (2005) Cryo-electron microscopy reconstruction of a poliovirus-receptor-membrane complex. *Nat Struct Mol Biol* **12**: 615–618.
119. Kirkegaard K. (1990) Mutations in VP1 of poliovirus specifically affect both encapsidation and release of viral RNA. *J Virol* **64**: 195–206.
120. Moscufo N, Yafal A, Rogove A, *et al.* (1993) A mutation in VP4 defines a new step in the late stages of cell entry by poliovirus. *J Virol* **67**: 5075–5078.
121. Prchla E, Plank C, Wagner E, *et al.* (1995) Virus-mediated release of endosomal content *in vitro*: different behavior of adenovirus and rhinovirus serotype 2. *J Cell Biol* **131**: 111–123.
122. Schober D, Bayer N, Murphy RF, *et al.* (1999) Establishment of an assay to determine adenovirus-induced endosome rupture required for receptor-mediated gene delivery. *Gene Ther Mol Biol* **3**: 25–33.

123. Parente RA, Nir S, Szoka FCJ. (1990) Mechanism of leakage of phospholipid vesicle contents induced by the peptide GALA. *Biochemistry* **29**: 8720–8728.
124. Kienberger F, Zhu R, Moser R, *et al.* (2004) Monitoring RNA release from human rhinovirus by dynamic force microscopy. *J Virol* **78**: 3203–3209.
125. Kuznetsov YG, Daijogo S, Zhou J, *et al.* (2005) Atomic force microscopy analysis of icosahedral virus RNA. *J Mol Biol* **347**: 41–52.
126. Winther B, Gwaltney JM, Jr., Mygind N, *et al.* (1986) Sites of rhinovirus recovery after point inoculation of the upper airway. *JAMA* **256**: 1763–1767.
127. Hayden FG. (2004) Rhinovirus and the lower respiratory tract. *Rev Med Virol* **14**: 17–31.
128. Mosser AG, Brockman-Schneider R, Amineva S, *et al.* (2002) Similar frequency of rhinovirus-infectible cells in upper and lower airway epithelium. *J Infect Dis* **185**: 734–743.
129. Pitkaranta A, Hayden FG. (1998) Rhinoviruses: important respiratory pathogens. *Ann Med* **30**: 529–537.
130. Pitkaranta A, Puhakka T, Makela MJ, *et al.* (2003) Detection of rhinovirus RNA in middle turbinate of patients with common colds by *in situ* hybridization. *J Med Virol* **70**: 319–323.
131. Knight DA, Holgate ST. (2003) The airway epithelium: structural and functional properties in health and disease. *Respirology* **8**: 432–446.
132. Demoly P, Sahla M, Campbell AM, *et al.* (1998) ICAM-1 expression in upper respiratory mucosa is differentially related to eosinophil and neutrophil inflammation according to the allergic status. *Clin Exp Allergy* **28**: 731–738.
133. Papon JF, Coste A, Gendron MC, *et al.* (2002) HLA-DR and ICAM-1 expression and modulation in epithelial cells from nasal polyps. *Laryngoscope* **112**: 2067–2075.
134. Winther B, Greve JM, Gwaltney JM, *et al.* (1997) Surface expression of intercellular adhesion molecule 1 on epithelial cells in the human adenoid. *J Infect Diseases* **176**: 523–525.
135. Taguchi M, Sampath D, Koga T, *et al.* (1998) Patterns for RANTES secretion and intercellular adhesion molecule 1 expression mediate transepithelial T cell traffic based on analyses *in vitro* and *in vivo*. *J Exp Med* **187**: 1927–1940.

136. Lopez-Souza N, Dolganov G, Dubin R, *et al.* (2004) Resistance of differentiated human airway epithelium to infection by rhinovirus. *Am J Physiol Lung Cell Mol Physiol* **286**: L373–381.
137. Suzuki T, Yamaya M, Kamanaka M, *et al.* (2001) Type 2 rhinovirus infection of cultured human tracheal epithelial cells: role of LDL receptor. *Am J Physiol Lung Cell Mol Physiol* **280**: L409–420.
138. Suzuki T, Yamaya M, Sekizawa K, *et al.* (2001) Bafilomycin A(1) inhibits rhinovirus infection in human airway epithelium: effects on endosome and ICAM-1. *Am J Physiol Lung Cell Mol Physiol* **280**: L1115–1127.

This page intentionally left blank

Chapter 2

Role of Lipid Microdomains in Influenza Virus Multiplication

*Makoto Takeda**,†

Lipid bilayers of biological membranes have often been considered as two-dimensional solvents of integral membrane proteins, and membrane lipids have been thought to have minor functions for membrane phenomena (fluid-mosaic model). A current concept of lipid rafts refined this classical model, proposing that membrane lipids play key roles for regulation of membrane functions — such as membrane trafficking and signal transduction — by forming functional lipid microdomains termed rafts where specific membrane proteins and lipids become incorporated. Viruses are obligate parasites of living organisms and their replication is absolutely dependent on the host cell machinery. Many viruses have been shown to use membrane rafts during one or more steps of their replication cycles. Influenza virus represents a well-studied example for the interactions of viruses and rafts. Intrinsic raft-associating properties of the influenza virus envelope glycoproteins were shown to promote virus replication by facilitating virus assembly. Additional functions of rafts in membrane fusion, glycoprotein maturation, and transport have also been suggested.

Introduction

The fluid mosaic model by Singer and Nicolson proposed that biological membranes are analogous to a two-dimensional oriented solution of integral membrane proteins in the viscous phospholipid bilayer

*Corresponding author.
†Department of Virology, Faculty of Medicine, Kyushu University, Fukuoka 812-8582, Japan. Tel: +81 92 642 6138 Fax: +81 92 642 6140; E-mail: mtakeda@virology.med.kyushu-u.ac.jp.

solvent.[1] Within this model, membrane lipids have been thought to distribute mostly homogeneously and play minor functions in membrane phenomena. A current concept of lipid rafts refined this classical model, proposing instead that the membrane lipids actively participate in the regulation of membrane functions by forming laterally organized microdomains that selectively incorporate or exclude membrane proteins.[2] This concept has had a great impact on studies of membrane trafficking and signal transduction. It also provided a new view to allow understanding of how viruses regulate their viral proteins by specifying interactions with membrane lipids to promote their replication cycles.

Membrane Lipids Form Functional Microdomains

Biological cell membranes are composed primarily of phospholipids (sphingomyelin and glycerophospholipids), glycolipids (mostly glycosphingolipids), and cholesterol (Fig. 1A). Phospholipids and glycolipids have two acyl-chain tails. Membrane fluidity is affected by the degree of saturation and by the length of these acyl-chains. Since both acyl-chains of sphingolipids (sphingomyelin and glycosphingolipids) are usually saturated and longer than those of glycerophospholipids (phosphatidylethanolamine, phosphatidylserine, and phosphatidylcholine), these lipids tend to pack more tightly against one another than glycerophospholipids. On the other hand, one of the acyl-chains of glycerophospholipids is usually kinked by *cis*-double bonds (unsaturated), causing glycerophospholipids to have weak interactions with one another. As a result, membrane lipids of bilayers composed predominantly of glycerophospholipids are disordered and highly fluid; that is, they reside in a liquid-disordered, or liquid-crystalline (lc), phase (Fig. 1B). Cholesterol molecules take their positions between phospholipids and glycolipids, strengthening the order of these lipids. With cholesterol, sphingolipids become highly ordered without losing their rotational and lateral mobility, and reside in a liquid-ordered (lo) phase[3] (Fig. 1B). Since lipid molecules in lo phase are insoluble in solutions of nonionic detergents at low temperature (such as 0.25–1.0% Triton X100 at 4°C), it is possible to isolate them from low-density fractions after floatation in sucrose gradients.[4] Using this method, it has been shown that biologcal membranes contain cholesterol and sphingolipid-rich

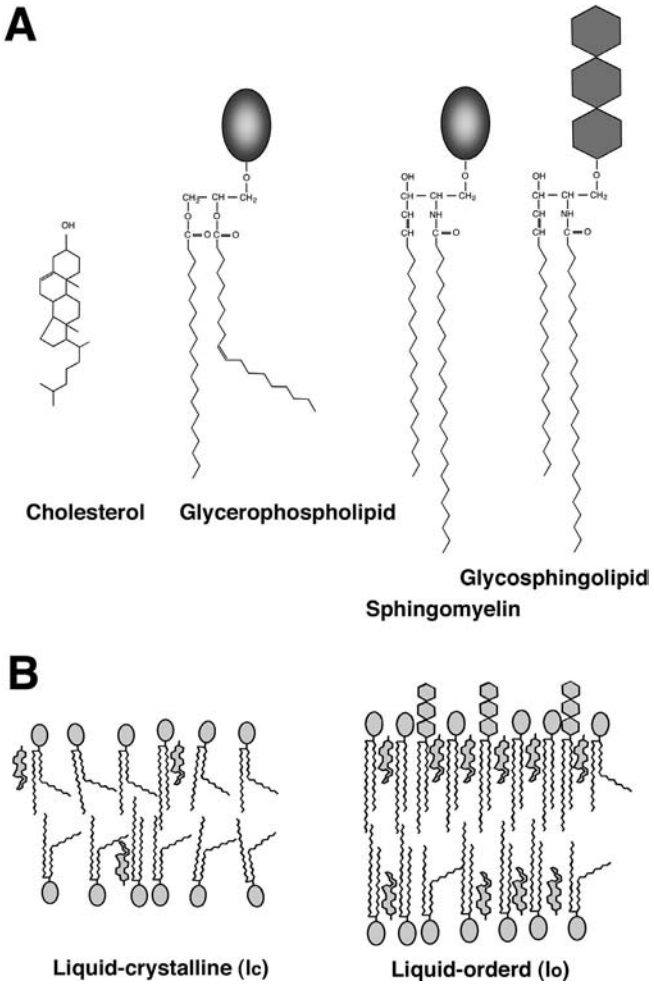


Fig. 1. Lipid molecules of biological membranes and phases of the membrane lipids. (**A**) The major classes of biological membrane lipids are phospholipids (glycerophospholipids and sphingomyelin), glycosphingolipids, and cholesterol. Acyl chains of sphingolipids (sphingomyelin and glycosphingolipids) are saturated, whereas an acyl chain of glycerophospholipids is usually kinked by a *cis* double bond (unsaturated). (**B**) Lipid bilayers composed predominantly of glycerophospholipids are highly fluid as proposed by the fluid mosaic model,[1] where lipid molecules reside in a liquid-crystalline (lc) (or liquid-disordered) phase. With cholesterol, sphingolipids become highly ordered but retain their lateral and rotational mobility, residing in a liquid-ordered (lo) phase. Lipid molecules of raft microdomains are thought to be in a lo or a lo-like phase.

microdomains, where lipid molecules reside in lo phase or lo-like phase.[3] Simons and Ikonen showed that these microdomains incorporate specific membrane proteins, such as glycosyl-phosphatidyl-inositol (GPI)-anchored proteins and doubly acylated tyrosine kinases of the Src family, and that they play crucial roles in membrane functions, such as membrane trafficking and signal transduction.[2] It was postulated that these microdomains move around the plane of cell membranes accommodating selected membrane proteins and that they function as communication centers of membrane proteins; hence, these microdomains were termed as moving platforms or functional rafts.[2] Because of the dynamic properties and the small size of the microdomains, it has been difficult to obtain direct evidence of their existence in living cells. However, data from a fluorescence resonance energy transfer technique and a cross-linking method strongly suggested that lipid microdomains do exist in living cells.[5–7] When studying these raft microdomains, it should be kept in mind that they are dynamic structures that continuously change in terms of size, shape, and proteins incorporated.

Lipid Rafts and Viruses

Viruses are obligate parasites of living organisms and their replication is absolutely dependent on the host cell machinery. It is now widely accepted that membrane lipids can control protein transport and regulate interaction of intracellular, extracellular, and membrane proteins by forming dynamic lipid microdomains. These properties of membrane lipids are thought to be fundamental for eukaryotic cells, and many cellular and viral membrane proteins appear to serve the function of specifying interactions with membrane lipids. Viruses that have been shown to use lipid microdomains are listed in Table 1. The number of such viruses is increasing, and ultimately most enveloped viruses might be included in this list.[8–12] It will be important to learn how these viruses use lipid microdomains in their replication cycles. Influenza A virus (influenza virus) is the best-studied example of the use of lipid rafts by a virus. In this review we discuss how influenza virus uses lipid microdomain rafts to promote its replication cycles.

Table 1. Raft-dependent Processes and Virus

Entry	
Human immunodeficiency virus 1	Manes *et al.*, 2000; Popik *et al.*, 2002
Human T-cell leukemia virus 1	Niyogi and Hildreth, 2001
Murine leukemia virus	Lu and Silver 2000
Human coronavirus 229E	Nomura *et al.*, 2004
Echovirus 11	Stuart *et al.*, 2002
Echovirus 1	Pietiainen *et al.*, 2004
Coxsackievirus A9	Triantafilou and Triantafilou, 2003
Coxsackievirus B4	Triantafilou and Triantafilou, 2004
Rotavirus	Isa *et al.*, 2004
Herpes simplex virus	Bender *et al.*, 2003
Epstein-Barr virus	Katzman and Longnecker, 2003
Assembly/budding	
Human immunodeficiency virus 1	Nguyen and Hildreth, 2000; Ono and Freed, 2001; Lindwasser and Resh, 2001;
Human T-cell leukemia virus 1	Feng *et al.*, 2003
Influenza A virus	Scheiffele *et al.*, 1999; Zhang *et al.*, 2000 Takeda *et al.*, 2003
Measles virus	Manie *et al.*, 2000; Vincent *et al.*, 2000
Sendai virus	Ali and Nayak, 2000
Respiratory syncytial virus	Brown *et al.*, 2002; Henderson *et al.*, 2002
Ebola virus	Bavari *et al.*, 2002; Panchal *et al.*, 2003
Marburg virus	Bavari *et al.*, 2002
Rotavirus	Sapin *et al.*, 2002
Herpes simplex virus	Lee *et al.*, 1999
Hepatitis C virus	Matto *et al.*, 2004
RNA synthesis	
Respiratory syncytial virus	McDonald *et al.*, 2004
Hepatitis C virus	Shi *et al.*, 2003; Aizaki *et al.*, 2004

Influenza Virus Glycoproteins Intrinsically Associate with Rafts

Influenza A virus is an enveloped virus that belongs to the family *Orthomyxoviridae*. The virus genome is comprised of eight segments of negative-sense RNAs. Eleven proteins are encoded in the genome.[13] Of these, four proteins are thought to be membrane proteins: three

transmembrane proteins, namely hemagglutinin (HA), neuraminidase (NA), and M2 proteins; and a peripheral membrane protein, namely matrix (M1) protein (Fig. 2). HA and NA can be seen by electron microscopy as protruding spikes of the virus envelope. Influenza virus attaches to cells as a result of binding by HA to sialic-acid-containing receptors and enters into cells by receptor-mediated endocytosis. In the acidic environment of endosomes, HA undergoes conformational changes that trigger fusion between virus envelopes and endosomal membranes. Conversely, NA facilitates virus release from cells by removing sialic acids from cells and progeny viruses.[13] HA and NA have been shown to associate intrinsically with rafts, and residues within the transmembrane (TM) domains of these proteins are important for their raft-association.[14–17] Mutations at the exoplasmic half of the HA and NA TM domains severely impaired raft association.[14–18] Cytoplasmic tails (CTs) of these proteins also contribute to their raft-association.[19] In addition, HA requires three palmitoylated cystein residues to associate with rafts.[19,20] Of these three cysteins, one is in the TM domain and the others are in the CT. All three are highly conserved among 15 HA subtypes.[21] Influenza virus glycoproteins are therefore thought to have evolved to associate with rafts. Why do influenza virus glycoproteins need to associate with rafts? It has been shown that in polarized cells influenza virus buds exclusively from apical cell membranes,[22] and analyses of lipid composition of purified virions have suggested that influenza virus assembles at and buds from rafts.[19,23] Both HA and NA are transported to the apical cell surface when expressed alone,[24,25] and cholesterol was needed for HA apical transport.[26] TM domains of both HA and NA were shown to contain apical sorting signals.[18,27] In light of this, an attractive speculation was that influenza virus uses rafts as a carrier to target its glycoproteins to the apical cell surface, where virus particles assemble and bud. In support of this idea, some mutations that abolished the raft-association of HA or NA retargeted these glycoproteins to the basolateral membrane surface.[14–16,18] However, other mutagenesis results indicate that raft-association does not necessarily correlate with the apical transport of HA and NA.[14,15,19,27] A mutant HA that almost completely lost the ability to associate with rafts was still transported to the apical cell surface.[17]

These data therefore suggest that raft-association is not obligatory for the apical transport of influenza virus glycoproteins and that influenza virus can use other cellular machinery or microdomains to transport HA and NA to the apical cell surface. Another raft-like lipid microdomain may regulate the transport of influenza virus glycoproteins.

M1 is a peripheral membrane protein and plays a central role in virus assembly and budding.[28] M1 is most abundant in virions and thought to interact with both inside ribonucleoprotein (RNP) cores and outside envelope glycoproteins, HA and NA[29–31] (Fig. 2A). When M1 was expressed alone, only a small amount of M1 associated with rafts.[17,19,30] However, the amount of M1 isolated from raft membrane fractions increased significantly by co-expressing HA and NA.[17,19,29,30] M2 is the third integral membrane protein and functions as a proton ion channel required for virus uncoating.[13] M2 is transported selectively to apical cell membranes, like HA and NA.[32] However, no significant association with rafts was shown, unlike HA and NA.[17,19] These data may explain inefficient incorporation of M2 into virions.[19,33] Recently, Carrasco *et al.* showed that nucleoprotein (NP), the major component of RNP, was also transported to apical cell membranes and that this apical transport was raft-dependent.[34]

Raft-Association of Viral Glycoproteins Promotes Influenza Virus Replication

As stated above, influenza virus HA has three palmitoylated cystein residues. These cysteins were shown to be needed for HA raft-association. However, a recombinant influenza virus having a mutated HA, where the three cysteins were substituted with other residues, could be generated by a revere genetics method and was shown to be viable.[35] This observation may suggest that the raft-associating property of HA is not so important for influenza virus replication. To address this issue, nine recombinant influenza viruses have been generated whose HA proteins contained a series of alanine substitutions at their TM domains.[17] All mutated HAs were transported to and expressed on the cell surface as efficiently as the wild-type (wt) HA, but they differed in their ability to associate with rafts. Some of them

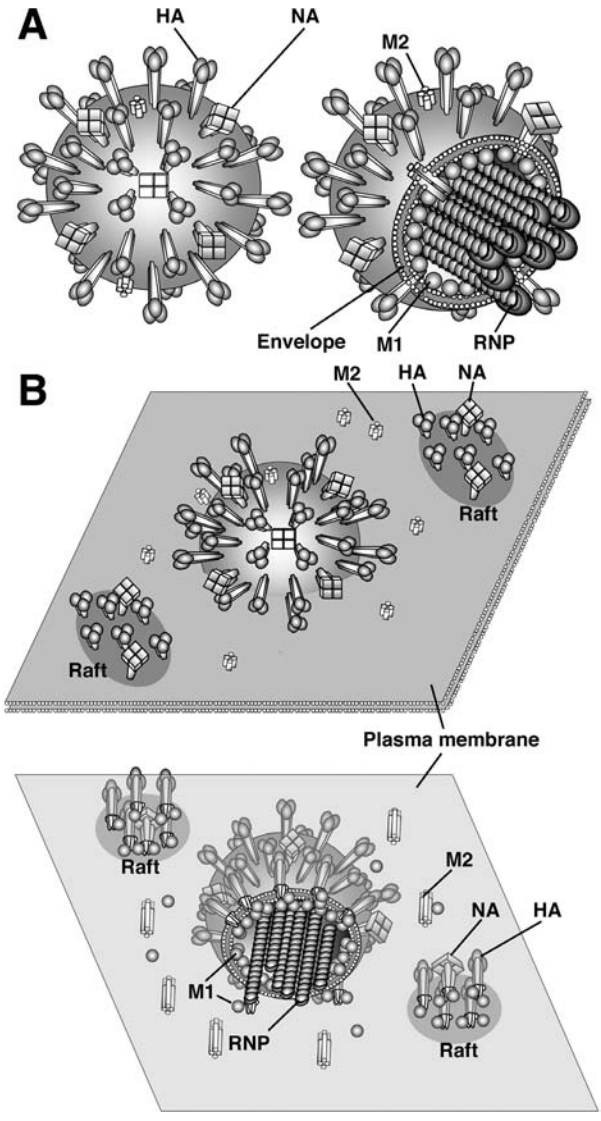


Fig. 2. Structure of influenza virus and a model of virus assembly. (**A**) Influenza virus is an enveloped virus with two major envelope glycoprotein spikes (hemagglutinin [HA] and neuraminidase [NA]). Each virion contains several M2 ion channel tetramers on the envelope. M1 is a peripheral membrane protein and thought to bridge envelope glycoproteins and ribonucleoprotein (RNP) cores, playing a central role in virus assembly. A virion of the right side is partially cut to illustrate the inside of the virion.

almost completely lost the ability to associate with rafts (nonraft HA). The mutant influenza viruses that had nonraft HA were shown to be morphologically indistinguishable from wt virus by electron microscopy, but importantly they contained smaller amounts of HA and replicated poorly in cells.[17] These data showed that HA needs to associate with rafts for influenza virus to replicate efficiently. Barman *et al.* (2004) performed a similar experiment with NA.[36] Recombinant influenza viruses were generated in which NA contained a series of alanine substitutions in the TM domain and CT, or the TM domain of NA was replaced with that of a transferin receptor.[36] These mutations within the TM domain and CT had multiple effects on NA function and virus replication. They affected raft-association, expression, transport, and enzyme activity of NA. Reduced incorporation of NA and an altered morphology of budding virions were also caused by the NA mutations.[36] Each NA mutant virus had a different set of defects. However, generally speaking, mutations of the ectodomain-proximal amino acids of the TM domain progressively lowered affinity of NA with rafts, reduced NA enzyme activity, and attenuated virus growth.[36] Because of the multiple effects of NA mutation, it was unclear which defect was directly due to the lowered affinity with rafts. However, it is likely that NA of influenza virus needs to associate with rafts for proper maturation, transport, and efficient incorporation into virions, to allow efficient virus replication.

HA Fusion and Lipid Rafts

How does HA promote virus replication by interacting with rafts? HA functions as homotrimers and is responsible for virus binding and subsequent membrane fusion. HA is synthesized as a single polypeptide, HA_0, and is subsequently cleaved by trypsin-like cellular proteases

(**B**) At the plasma membranes, HA and NA concentrate in raft microdomains. M1 is also accumulated in rafts by interacting with cytoplasmic tails of HA and NA. On the other hand, M2 is largely excluded from rafts. RNP cores associate with rafts intrinsically or by binding with M1, promoting the assembly processes of budding virions. Views from the outside (*top*) and from the inside (*bottom*) of the cells are shown.

into two disulfide-linked subunits, HA_1 and HA_2. This cleavage is a prerequisite for HA to undergo conformational change to a fusion-competent form in response to the low pH environment of endosomes. HA trimers have been shown to function cooperatively during the process of membrane fusion.[37–39] The minimum number of HA trimers required to initiate a fusion event was estimated to be three to four.[38] Many enveloped viruses, including influenza virus, human immunodeficiency virus (HIV), Ebola virus, and paramyxoviruses, share a very similar molecular mechanism for membrane fusion.[40] For HIV, it was also estimated that cooperation of three to six Env trimers was required for formation of a fusion pore.[41] If these trimers of fusion proteins (such as HA and Env) distribute randomly on a vast field of plasma membranes, how efficiently can they find their cooperation partners for the process of membrane fusion? It has been established that HA expressed on the plasma membrane by virus infection distributes in clusters at the sites of virus budding.[42] Recently we showed that rafts have a crucial role for this clustering of HA.[17] We found that nonraft HA mostly distributed homogeneously on the plasma membrane.[17] We compared fusion activity of the nonraft HA expressed on the cell surface with that of wt HA (cell-to-cell fusion) by lipid mixing and aqueous content mixing assays (HA-expressing BHK cells and dual-labeled human erythrocytes were used as effecter cells and target cells, respectively).[17] At any expression level, either wt or nonraft HA caused complete fusion, but the number of fusion events by nonraft HA was reduced (~60%).[17] This reduced fusion activity may be due to the lowered local density of HA. Similarly, nonraft HA virus showed a reduced virus-to-cell fusion activity in a fluorescent dequenching fusion assay with R18 (lipid fluorescence probe)-labeled virions and red blood cell ghosts (target membranes).[17] Since nonraft HA virions contained a smaller amount of HA than wt virions (~60%), the reduced fusion activity by nonraft HA virus could also be explained by the lowered density of HA. However, other mechanisms can be proposed, such as a need for HA trimers to be stabilized by lipid molecules in rafts for efficient membrane fusion, or a need for HA trimers to form a higher-order structure with lipids in rafts. When cholesterol was depleted from purified virions by methyl-β-cyclodextrin (MβCD), virus infectivity was reduced by 10-fold to 100-fold.[17,43]

These data suggest a structural requirement of rafts on virus envelopes for efficient virus entry. However, in spite of these strong reductions of infectivity, we found no defect in virus-to-cell fusion activity of MβCD-treated virus by a fluorescent dequenching fusion assay.[17] One possibility is that membrane fusion between virus envelopes and cell membranes may be interrupted at the hemifusion stage; that is, lipid mixing occurred, but subsequent content mixing did not. Due to the difficulty of direct measurement of the content mixing between virus and cells, we performed a hemolysis assay.[17] Again, MβCD-treated viruses showed no defect on hemolysis. Reduced infectivity of MβCD-treated viruses may be at least partially explained by an effect on M2 protein, since the uncoating process supported by the M2 ion channel function is critical for influenza virus infectivity and cholesterol was shown to be required for the M2 ion channel activity.[44,45] Further insights into the requirement for cholesterol or lipid microdomains on virus envelopes will require further investigation.

Rafts Function at the Process of Influenza Virus Assembly

The process of virus assembly includes recruitment of all structural viral components at a specific site and segregation of cellular proteins from the structural viral components. The molecular mechanism of this process has been poorly understood. The concept of lipid rafts functioning during assembly as selection and concentration devices of specific membrane proteins is therefore an attractive one. Many viruses, including influenza virus, have been shown to use rafts as the sites of assembly (Table 1).[8,9,11,12] The lipid composition of purified influenza viruses is rich in cholesterol and sphingolipids, similar to that of the membrane rafts.[19,23] These data are consistent with the observations that both envelope glycoproteins HA and NA associate with rafts. In addition to the TM domains, CTs of HA and NA have been shown to play an important role in the raft-association and in virus assembly, most likely by interacting with M1 proteins.[19,46] Particle production of a mutant influenza virus with CT deletions of both HA and NA was reduced ~10-fold.[46] These particles are greatly elongated or distended[46] and contain a reduced amount of cholesterol

and sphingolipids.[19] When affinity of HA or NA with rafts was lowered by mutagenesis, the virus contained a reduced amount of HA and NA, respectively.[17,36] A decrease in the number of HA spikes in a nonraft HA virus was compensated for by an increase in the number of NA spikes that associated with rafts.[17] Considering these data together, we can illustrate the assembly process of influenza virus as follows (Fig. 2B). Envelope glycoproteins associate with rafts at the Golgi apparatus.[17,47] This association seems to be required for NA to gain maximum enzyme activity and be efficiently transported to the cell surface.[36] During cell surface transport, both HA and NA are concentrated in the microdomain rafts, whereas M2 ion channel proteins are largely excluded from the rafts.[17,19] At the plasma membrane, CTs of HA and NA concentrated in the microdomain rafts function as docks for M1,[29,30] which is the major driving force of virus budding.[28] Consequently, these raft regions act as the assembly sites of budding viruses.[23] M1 proteins are thought to interact with both CTs of glycoproteins and NP, a major constituent of RNP cores. M1 can thus drag RNP cores into the assembly sites. The intrinsic raft-associating property of NP may also contribute to transport of RNPs to assembly sites.[34] Dragged RNP cores further promote the budding process. Finally, the budding viruses are enveloped with the raft-containing membranes with tightly packed viral glycoproteins that allow the viruses to initiate the next round of infections efficiently.[17]

Concluding Remarks

Lipid molecules of biological membranes have been thought to function as a simple solvent of membrane proteins. Now, it is recognized that these molecules form multi-functional microdomains. The possibility that these microdomains function in an important way to allow concentration of selected biological components within membranes has made them an attractive focus of study for many cell biologists, immunologists, and virologists. Lipid microdomain rafts are generally isolated as detergent-insoluble membrane fractions. However, direct evidence of the existence of these microdomains in living cells has been difficult to obtain, raising some skepticism as to their existence.[48]

The main argument of the raft concept is that membrane lipids form dynamic microdomains that play key roles in membrane functions. Increasing evidence supports this concept of lipid rafts. Several studies have proposed the existence of other lipid microdomains with different properties from the classical rafts.[49–51] Lines of evidence have been accumulating that many viruses (probably most enveloped viruses) replicate in a way that depends on the functions of these lipid microdomains. A better understanding of the behavior of these microdomains may provide new insights that are generally applicable for virus replication in cells, and offer the possibility of new strategies for counteracting the artful viruses.

Acknowledgments

I thank Dr. R. A. Lamb for kind suggestion. I also thank Drs. Y. Yanagi and A. P. Schmitt for their critical reading and helpful comments.

References

1. Singer SJ, Nicolson GL. (1972) The fluid mosaic model of the structure of cell membranes. *Science* **175**: 720–731.
2. Simons K, Ikonen E. (1997) Functional rafts in cell membranes. *Nature* **387**: 569–572.
3. Brown DA, London E. (1998) Functions of lipid rafts in biological membranes. *Annu Rev Cell Dev Biol* **14**: 111–136.
4. Brown DA, Rose JK. (1992) Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. *Cell* **68**: 533–544.
5. Pralle A, Keller P, Florin EL, *et al.* (2000) Sphingolipid-cholesterol rafts diffuse as small entities in the plasma membrane of mammalian cells. *J Cell Biol* **148**: 997–1008.
6. Varma R, Mayor S. (1998) GPI-anchored proteins are organized in submicron domains at the cell surface. *Nature* **394**: 798–801.
7. Friedrichson T, Kurzchalia TV. (1998) Microdomains of GPI-anchored proteins in living cells revealed by crosslinking. *Nature* **394**: 802–805.
8. Briggs JA, Wilk T, Fuller SD. (2003) Do lipid rafts mediate virus assembly and pseudotyping? *J Gen Virol* **84**: 757–768.

9. Chazal N, Gerlier D. (2003) Virus entry, assembly, budding, and membrane rafts. *Microbiol Mol Biol Rev* **67**: 226–237.
10. Duncan MJ, Shin JS, Abraham SN. (2002) Microbial entry through caveolae: variations on a theme. *Cell Microbiol* **4**: 783–791.
11. Manes S, del Real G, Martinez-A C. (2003) Pathogens: raft hijackers. *Nat Rev Immunol* **3**: 557–568.
12. Suomalainen M. (2002) Lipid rafts and assembly of enveloped viruses. *Traffic* **3**: 705–709.
13. Lamb RA, Krug RM. (2001) Orthomyxoviridae: the viruses and their replication. In D.M. Knipe & P.M. Howlwy (eds), *Fields virology (4th ed.)*, pp. 1487–1531. Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania.
14. Barman S, Nayak DP. (2000) Analysis of the transmembrane domain of influenza virus neuraminidase, a type II transmembrane glycoprotein, for apical sorting and raft association. *J Virol* **74**: 6538–6545.
15. Barman S, Ali A, Hui EK, *et al.* (2001) Transport of viral proteins to the apical membranes and interaction of matrix protein with glycoproteins in the assembly of influenza viruses. *Virus Res* **77**: 61–69.
16. Scheiffele P, Roth MG, Simons K. (1997) Interaction of influenza virus haemagglutinin with sphingolipid-cholesterol membrane domains via its transmembrane domain. *EMBO J* **16**: 5501–5508.
17. Takeda M, Leser GP, Russell CJ, Lamb RA. (2003) Influenza virus hemagglutinin concentrates in lipid raft microdomains for efficient viral fusion. *Proc Natl Acad Sci USA* **100**: 14610–14617.
18. Kundu A, Avalos RT, Sanderson CM, Nayak DP. (1996) Transmembrane domain of influenza virus neuraminidase, a type II protein, possesses an apical sorting signal in polarized MDCK cells. *J Virol* **70**: 6508–6515.
19. Zhang J, Pekosz A, Lamb RA. (2000) Influenza virus assembly and lipid raft microdomains: a role for the cytoplasmic tails of the spike glycoproteins. *J Virol* **74**: 4634–4644.
20. Melkonian KA, Ostermeyer AG, Chen JZ, *et al.* (1999) Role of lipid modifications in targeting proteins to detergent-resistant membrane rafts. Many raft proteins are acylated, while few are prenylated. *J Biol Chem* **274**: 3910–3917. Adv Virus Res. 58:1–28, 2002.
21. Nobusawa E, Aoyama T, Kato H, *et al.* (1991) Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza A viruses. *Virology* **182**: 475–485.

22. Boulan ER, Sabatini DD. (1978) Asymmetric budding of viruses in epithelial monlayers: a model system for study of epithelial polarity. *Proc Natl Acad Sci USA* **75**: 5071–5075.
23. Scheiffele P, Rietveld A, Wilk T, Simons K. (1999) Influenza viruses select ordered lipid domains during budding from the plasma membrane. *J Biol Chem* **274**: 2038–2044.
24. Roth MG, Compans RW, Giusti L, *et al.* (1983) Influenza virus hemagglutinin expression is polarized in cells infected with recombinant SV40 viruses carrying cloned hemagglutinin DNA. *Cell* **33**: 435–443.
25. Jones LV, Compans RW, Davis AR, *et al.* (1985) Surface expression of influenza virus neuraminidase, an amino-terminally anchored viral membrane glycoprotein, in polarized epithelial cells. *Mol Cell Biol* **5**: 2181–2189.
26. Keller P, Simons K. (1998) Cholesterol is required for surface transport of influenza virus hemagglutinin. *J Cell Biol* **140**: 1357–1367.
27. Lin S, Naim HY, Rodriguez AC, Roth MG. (1998) Mutations in the middle of the transmembrane domain reverse the polarity of transport of the influenza virus hemagglutinin in MDCK epithelial cells. *J Cell Biol* **142**: 51–57.
28. Gómez-Puertas P, Albo C, Perez-Pastrana E, *et al.* (2000) Influenza virus matrix protein is the major driving force in virus budding. *J Virol* **74**: 11538–11547.
29. Ali A, Nayak DP. (2000) Assembly of Sendai virus: M protein interacts with F and HN proteins and with the cytoplasmic tail and transmembrane domain of F protein. *Virology* **276**: 289–303.
30. Ali A, Avalos RT, Ponimaskin E, Nayak DP. (2000) Influenza virus assembly: effect of influenza virus glycoproteins on the membrane association of M1 protein. *J Virol* **74**: 8709–8719.
31. Enami M, Enami K. (1966) Influenza virus hemagglutinin and neuraminidase glycoproteins stimulate the membrane association of the matrix protein. *J Virol* **70**: 6653–6657.
32. Hughey PG, Compans RW, Zebedee SL, Lamb RA. (1992) Expression of the influenza A virus M2 protein is restricted to apical surfaces of polarized epithelial cells. *J Virol* **66**: 5542–5552.
33. Zebedee SL, Lamb RA. (1988) Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions. *J Virol* **62**: 2762–2772.

34. Carrasco M, Amorim MJ, Digard P. (2004) Lipid raft-dependent targeting of the influenza a virus nucleoprotein to the apical plasma membrane. *Traffic* **5**: 979–992.
35. Jin H, Subbarao K, Bagai S, *et al.* (1996) Palmitylation of the influenza virus hemagglutinin (H3) is not essential for virus assembly or infectivity. *J Virol* **70**: 1406–1414.
36. Barman S, Adhikary L, Chakrabarti AK, *et al.* (2004) Role of transmembrane domain and cytoplasmic tail amino acid sequences of influenza A virus neuraminidase in raft association and virus budding. *J Virol* **78**: 5258–5269.
37. Blumenthal R, Sarkar DP, Durell S, *et al.* (1996) Dilation of the influenza hemagglutinin fusion pore revealed by the kinetics of individual cell-cell fusion events. *J Cell Biol* **135**: 63–71.
38. Danieli T, Pelletier SL, Henis YI, White JM. (1996) Membrane fusion mediated by the influenza virus hemagglutinin requires the concerted action of at least three hemagglutinin trimers. *J Cell Biol* **133**: 559–569.
39. Markovic I, Leikina E, Zhukovsky M, *et al.* (2001) Synchronized activation and refolding of influenza hemagglutinin in multimeric fusion machines. *J Cell Biol* **155**: 833–844.
40. Lamb RA, Joshi SB, Dutch RE. (1999) The paramyxovirus fusion protein forms an extremely stable core trimer: structural parallels to influenza virus haemagglutinin and HIV-1 gp41. *Mol Membr Biol* **16**: 11–19.
41. Doms RW, Trono D. (2000) The plasma membrane as a combat zone in the HIV battlefield. *Genes Dev* **14**: 2677–2688.
42. Compans RW, Dimmock NJ. (1969) An electron microscopic study of single-cycle infection of chick embryo fibroblasts by influenza virus. *Virology* **39**: 499–515.
43. Sun X, Whittaker GR. (2003) Role for influenza virus envelope cholesterol in virus entry and infection. *J Virol* **77**: 12543–12551.
44. Cleverley DZ, Geller HM, Lenard J. (1997) Characterization of cholesterol-free insect cells infectible by baculoviruses: effects of cholesterol on VSV fusion and infectivity and on cytotoxicity induced by influenza M2 protein. *Exp Cell Res* **233**: 288–296.
45. Schroeder C, Heider H, Möncke-Buchner E, Lin TI. (2005) The influenza virus ion channel and maturation cofactor M2 is a cholesterol-binding protein. *Eur Biophys J* **34**: 52–66.
46. Jin H, Leser GP, Zhang J, Lamb RA. (1997) Influenza virus hemagglutinin and neuraminidase cytoplasmic tails control particle shape. *EMBO J* **16**: 1236–1247.

47. Fiedler K, Kobayashi T, Kurzchalia TV, Simons K. (1993) Glycosphingolipid-enriched, detergent-insoluble complexes in protein sorting in epithelial cells. *Biochemistry* **32**: 6365–6373.
48. Munro S. (2003) Lipid rafts: elusive or illusive? *Cell* **115**: 377–388.
49. Holm K, Weclewicz K, Hewson R, Suomalainen M. (2003) Human immunodeficiency virus type 1 assembly and lipid rafts: Pr55(gag) associates with membrane domains that are largely resistant to Brij98 but sensitive to Triton X-100. *J Virol* **77**: 4805–4817.
50. Röper K, Corbeil D, Huttner WB. (2000) Retention of prominin in microvilli reveals distinct cholesterol-based lipid microdomains in the apical plasma membrane. *Nat Cell Biol* **2**: 582–592.
51. Lindwasser OW, Resh MD. (2001) Multimerization of human immunodeficiency virus type 1 Gag promotes its localization to barges, raft-like membrane microdomains. *J Virol* **75**: 7913–7924.
52. Manes S, del Real G, Lacalle RA, *et al.* (2000) Membrane raft microdomains mediate lateral assemblies required for HIV-1 infection. *EMBO Rep* **1**: 190–196.
53. Popik W, Alce TM, Au WC. (2002) Human immunodeficiency virus type 1 uses lipid raft-colocalized CD4 and chemokine receptors for productive entry into CD4(+) T cells. *J Virol* **76**: 4709–4722.
54. Niyogi K, Hildreth JE. (2001) Characterization of new syncytium-inhibiting monoclonal antibodies implicates lipid rafts in human T-cell leukemia virus type 1 syncytium formation. *J Virol* **75**: 7351–7361.
55. Lu X, Silver J. (2000) Ecotropic murine leukemia virus receptor is physically associated with caveolin and membrane rafts. *Virology* **276**: 251–258.
56. Nomur R, Kiyota A, Suzaki E, *et al.* (2004) Human coronavirus 229E binds to CD13 in rafts and enters the cell through caveolae. *J Virol* **78**: 8701–8708.
57. Stuart AD, Eustace HE, McKee TA, Brown TD. (2002) A novel cell entry pathway for a DAF-using human enterovirus is dependent on lipid rafts. *J Virol* **76**: 9307–9322.
58. Pietiainen V, Marjomaki V, Upla P, *et al.* (2004) Echovirus 1 endocytosis into caveosomes requires lipid rafts, dynamin II, and signaling events. *Mol Biol Cell* **15**: 4911–4925.
59. Triantafilou K, Triantafilou M. (2003) Lipid raft microdomains: key sites for Coxsackievirus A9 infectious cycle. *Virology* **317**: 128–135.
60. Triantafilou K, Triantafilou M. (2004) Lipid-raft-dependent Coxsackievirus B4 internalization and rapid targeting to the Golgi. *Virology* **326**: 6–19.

61. Isa P, Realpe M, Romero P, *et al.* (2004) Rotavirus RRV associates with lipid membrane microdomains during cell entry. *Virology* **322**: 370–381.
62. Bender FC, Whitbeck JC, Ponce de Leon M, *et al.* (2003) Specific association of glycoprotein B with lipid rafts during herpes simplex virus entry. *J Virol* **77**: 9542–9552.
63. Katzman RB, Longnecker R. (2003) Cholesterol-dependent infection of Burkitt's lymphoma cell lines by Epstein-Barr virus. *J Gen Virol* **84**: 2987–2992.
64. Nguyen DH, Hildreth JE. (2000) Evidence for budding of human immunodeficiency virus type 1 selectively from glycolipid-enriched membrane lipid rafts. *J Virol* **74**: 3264–3272.
65. Ono A, Freed EO. (2001) Plasma membrane rafts play a critical role in HIV-1 assembly and release. *Proc Natl Acad Sci USA* **98**: 13925–13930.
66. Feng X, Heyden NV, Ratner L. (2003) Alpha interferon inhibits human T-cell leukemia virus type 1 assembly by preventing Gag interaction with rafts. *J Virol* **77**: 13389–13395.
67. Manie SN, Debreyne S, Vincent S, Gerlier D. (2000) Measles virus structural components are enriched into lipid raft microdomains: a potential cellular location for virus assembly. *J Virol* **74**: 305–311.
68. Vincent S, Gerlier D, Manie SN. (2000) Measles virus assembly within membrane rafts. *J Virol* **74**: 9911–9915.
69. Brown G, Rixon HW, Sugrue RJ. (2002) Respiratory syncytial virus assembly occurs in GM1-rich regions of the host-cell membrane and alters the cellular distribution of tyrosine phosphorylated caveolin-1. *J Gen Virol* **83**: 1841–1850.
70. Henderson G, Murray J, Yeo RP. (2002) Sorting of the respiratory syncytial virus matrix protein into detergent-resistant structures is dependent on cell-surface expression of the glycoproteins. *Virology* **300**: 244–254.
71. Bavari S, Bosio CM, Wiegand E, *et al.* (2002) Lipid raft microdomains: a gateway for compartmentalized trafficking of Ebola and Marburg viruses. *J Exp Med* **195**: 593–602.
72. Panchal RG, Ruthel G, Kenny TA, *et al.* (2003) *In vivo* oligomerization and raft localization of Ebola virus protein VP40 during vesicular budding. *Proc Natl Acad Sci USA* **100**: 15936–15941.
73. Sapin C, Colard O, Delmas O, *et al.* (2002) Rafts promote assembly and atypical targeting of a nonenveloped virus, rotavirus, in Caco-2 cells. *J Virol* **76**: 4591–4602.

74. Lee GE, Church GA, Wilson DW. (2003) A subpopulation of tegument protein vhs localizes to detergent-insoluble lipid rafts in herpes simplex virus-infected cells. *J Virol* **77**: 2038–2045.
75. Matto M, Rice CM, Aroeti B, Glenn JS. (2004) Hepatitis C virus core protein associates with detergent-resistant membranes distinct from classical plasma membrane rafts. *J Virol* **78**: 12047–12053.
76. McDonald TP, Pitt AR, Brown G, *et al.* (2004) Evidence that the respiratory syncytial virus polymerase complex associates with lipid rafts in virus-infected cells: a proteomic analysis. *Virology* **330**: 147–157.
77. Shi ST, Lee KJ, Aizaki H, *et al.* (2003) Hepatitis C virus RNA replication occurs on a detergent-resistant membrane that cofractionates with caveolin-2. *J Virol* **77**: 4160–4168.

This page intentionally left blank

Chapter 3

Functions of Integrin $\alpha 2\beta 1$, A Collagen Receptor, in the Internalization of Echovirus 1

Varpu Marjomäki, Mikko Huhtala†, Mikko Karjalainen*, Timo Hyypiä‡ and Jyrki Heino§*

$\alpha 2\beta 1$ is a collagen-binding integrin that has been found to act as a receptor for echovirus 1 (EV1), a picornavirus and a human pathogen. Integrin $\alpha 2\beta 1$ binds EV1 and the natural ligands with a special inserted domain (αI domain) in the integrin α subunit. Molecular modeling based on a cryo EM structure suggests that the exact binding sites are distinct but close enough to make simultaneous binding impossible. Binding of EV1 to the integrins causes a rapid clustering of the receptors on the plasma membrane, their internalization in complex with the virus and gradual accumulation into cytoplasmic caveosomes. The clustering is accompanied by rapid activation of protein kinase Cα, which is necessary for the internalization process. Confocal and electron microscopy have shown the presence of large complex tubulovesicular caveosomes containing both EV1 and $\alpha 2\beta 1$ integrin. The caveosomal nature of these structures has been verified by the accumulation of SV40 and EV1 in the same caveolin-1 enriched structures, and the lack of classical endosomal markers. Viral RNA may be released directly from the caveosomes to

*Department of Biological and Environmental Sciences, University of Jyväskylä, Survontie 9, FI-40500 Jyväskylä, Finland; E-mail: vmarjoma@cc.jyu.fi

†Department of Biochemistry, Åbo Akademi University, Tykistökatu 6A, FI-20520 Turku, Finland; E-mail: mikko.huhtala@abo.fi

‡Department Virology, University of Turku, Kiinamyllynkatu 13, FI-20520 Turku, Finland; E-mail: timo.hyypia@utu.fi

§Department of Biochemistry and Food Chemistry, University of Turku, Vatselankatu 2, FI-20014 Turku, Finland; E-mail: jyrki.heino@utu.fi

the cytosol, where it initiates the replication cycle. The molecular mechanisms of viral uncoating and release of RNA from caveosomes still remain unknown.

Integrin $\alpha 2\beta 1$ is the Cellular Receptor for Echovirus 1

The integrins are a large family of cell adhesion receptors. They are heterodimers formed by an α and a β subunit (for review see Ref. 1). The integrins that contain the αV subunit are internalization receptors for a large number of viruses, including adenoviruses, coxsackie viruses, foot-and-mouth-disease viruses and parechovirus 1.[2–6] αV integrins bind to a short RGD (arginine-glysine-aspartic acid) motif in both natural ligands and viruses, and they can guide the ligands into clathrin-coated vesicles and early endosomes.[2,7] In addition to αV integrins, other integrins may also act as virus receptors. Integrin $\alpha 2\beta 1$ belongs to the subgroup of collagen binding integrins (for review see Ref. 8). Echovirus 1 (EV1) binds to this integrin[9] and the virus-integrin complex is internalized via a different entry pathway when compared to αV integrin-dependent viruses.[10]

EV1 is a non-enveloped human pathogen belonging to the Enterovirus genus of the *Picornaviridae* family. In humans, echovirus infections are associated with meningitis, encephalitis, respiratory infections and diarrhea. EV1, like all picornavirus particles, is composed of a single RNA molecule of positive polarity and an icosahedral capsid consisting of VP1-4 proteins forming 60 protomers that surround the RNA. We have studied the molecular mechanisms of $\alpha 2\beta 1$ integrin–EV1 interactions to shed light on the role of receptor action in the process of virus entry.

EV1 Binds to α2I Domain in the $\alpha 2\beta 1$ Integrin and Induces Clustering of the Receptor Molecules

The integrins that act as collagen receptors contain a special inserted domain (αI domain) in their α subunit, which binds their natural ligands. The αI domain has a typical Rossman fold and a metal ion coordination site (MIDAS), containing an Mg^{++} ion. In α2I domain the metal ion and certain neighboring amino acid residues form

direct contacts with the natural ligands.[11] The natural ligands of α2β1 integrin include, in addition to various collagen subtypes, other matrix molecules such as laminin-1.[12] The α2I domain can recognize specific triple helical motifs in collagens and bind to them with a relatively high affinity. The best known motif is GFOGER (O = hydroxyproline).[13]

EV1 binds specifically to α2β1 integrin.[9] Blocking assays using monoclonal antibodies against putative EV1 receptors have shown that antibodies against α2 integrin subunits are the most potent blockers of EV1 infection.[9] Despite the fact that antibodies against β2 microglobulin have shown blocking effects, it is still unclear whether this molecule plays any direct role in EV1 infection.[9] Integrin α2β1 binds to EV1 using the α2I domain, and the binding site is close to the collagen-binding MIDAS. However, the two binding sites are clearly distinct and the adhesion of α2I to EV1 is not dependent on Mg^{++}.[14,15] Molecular modeling, based on the cryo-EM structure of α2I–EV1 complex, showed that α2I domain binds within the canyon (a depression surrounding the 5-fold symmetry axis) on the EV1 surface with contacts to both the inner and outer canyon walls.[16] Furthermore, the superposition of the α2I–collagen structure with the model of α2I–EV1 complex shows that when α2I is bound to EV1, the virus canyon floor makes it impossible for the collagens to bind.

EV1 may rely on the fact that on the plasma membrane some of the α2β1 integrins are usually free of natural ligands and therefore able to bind to the virus. However, in solid-phase binding assays the avidity of α2I domain to EV1 has been estimated to be at least 10-fold stronger than to collagen I.[16] The higher affinity of integrins to EV1 than to their natural ligand gives EV1 an opportunity to exploit this receptor for cellular entry even in collagenous tissues. In fact, our recent observations suggest that in cells inside 3D collagen gels there are receptors available to mediate viral infection (Marjomäki, unpublished).

The EV1 capsid is a symmetric structure that consists of 60 protomers, each containing one integrin binding site. The adjacent binding sites are located nearest to each other in the protomers around the 5-fold axis. Molecular modeling shows that multiple α2β1 integrin heterodimers, which are much larger structures that α2I domains alone, can simultaneously occupy adjacent binding sites on the EV1 capsid.[16] Furthermore, our observations with confocal microscopy have shown that treatment of

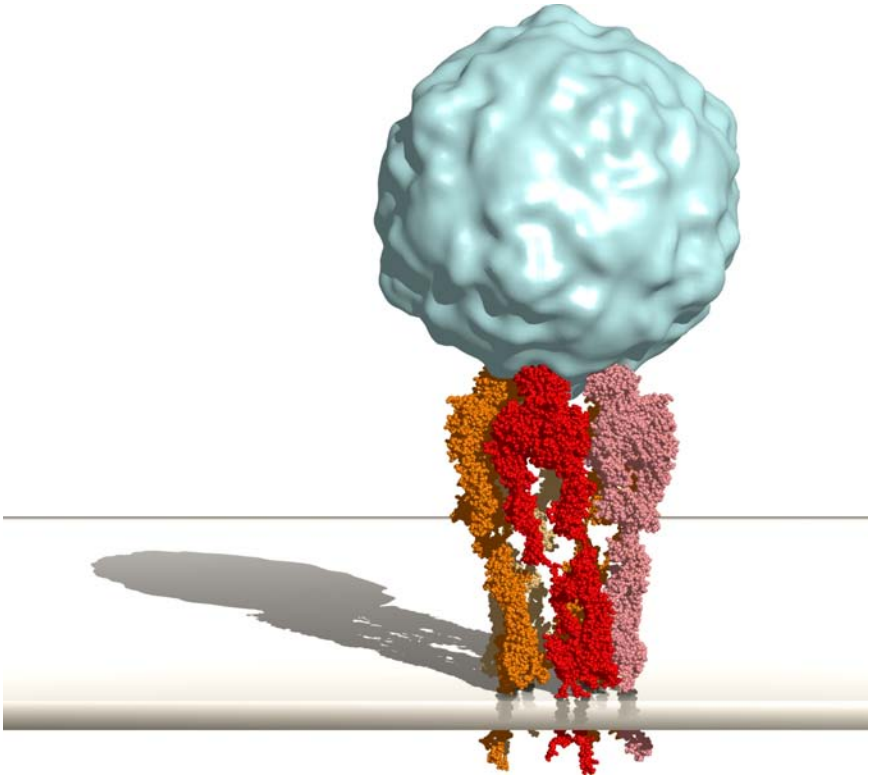


Fig. 1. Molecular model of EV1 with three $\alpha 2\beta 1$ integrins. EV1 is drawn as a cyan-colored isodensity surface of a map computed at 15 Å resolution from the high-resolution coordinates of the EV1 crystal structure (PDB entry 1EV1).[17] Three copies of a model of $\alpha 2\beta 1$ integrin heterodimer in the active conformation are drawn as space filling models in different shades of red and orange. The model of the integrin heterodimer is based on the crystal structures of $\alpha 2$I domain (1AOX),[18] $\alpha V\beta 3$ (1JV2)[19] and $\alpha IIb\beta 3$ (1TYE)[20] ectodomains, and published cryoelectron microscopy reconstructions. The binding site of the $\alpha 2$I domain on the surface has been solved by cryo-electron microscopy.[16] The orientation and position of the αI domain and the domains of the "stalk" region within the integrin heterodimer were modeled by hand and eye. Three copies of the integrin heterodimer fit without steric hindrace into the adjacent binding sites around the five-fold symmetry axis of the EV1 capsid. The image was made with PyMol 0.99 (DeLano Scientific, San Carlos, CA, USA) and POV-Ray 3.6 (www.povray.org).

cells with EV1 can cause a rapid clustering of the integrins on the plasma membrane.[16] The receptor clustering might be an important mechanism triggering integrin signaling and initiating virus entry (Fig. 1).

In contrast to many other viruses and their integrin receptors, binding of the α2I domain to EV1 does not directly induce uncoating of the virus.[16] Instead, the interaction seems to stabilize the virus. Therefore, it seems likely that the conditions needed for viral uncoating exist only inside cellular structures. Our sucrose gradient sedimentation analysis of the viral particles suggested that, while inside membrane vesicles, viral uncoating starts approximately 30 min post infection (p.i.), when approximately 25% of the virus particles are in the 80S form and represent capsids that have released their RNA genome.[9]

Clustering of α2β1 Integrin Leads to Rapid Internalization of the Virus-Receptor Complexes

We have recently tested the hypothesis that the integrin-mediated entry of EV1 is initiated by virus-dependent receptor clustering. For this purpose, we have used antibodies against α2β1 integrin and secondary antibodies to initiate clustering, and then followed living cells under a confocal microscope (Fig. 2.; Upla *et al.*, 2004; see http://www.molbiolcell.org/cgi/content/full/E03-08-0588/DC1 for the live cell videos). Immediately after the clusters formed they started a lateral movement along the cell surface. During the process smaller clusters seemed to have the tendency to fuse together. Based on the imaging of cells expressing actin-green fluorecent protein (GFP) chimeras, we propose that the integrin clusters may follow microfilaments. Already during the first 15 minutes after the formation of integrin clusters their internalization started. Thus the antibody-dependent integrin clusters were internalized in the same way as EV1, suggesting that receptor clustering is sufficient to initiate the entry process.

EV1 Internalization and Integrin Signaling

The binding of integrins to their ligands can induce both a conformational change in the receptors and receptor clustering. The integrins

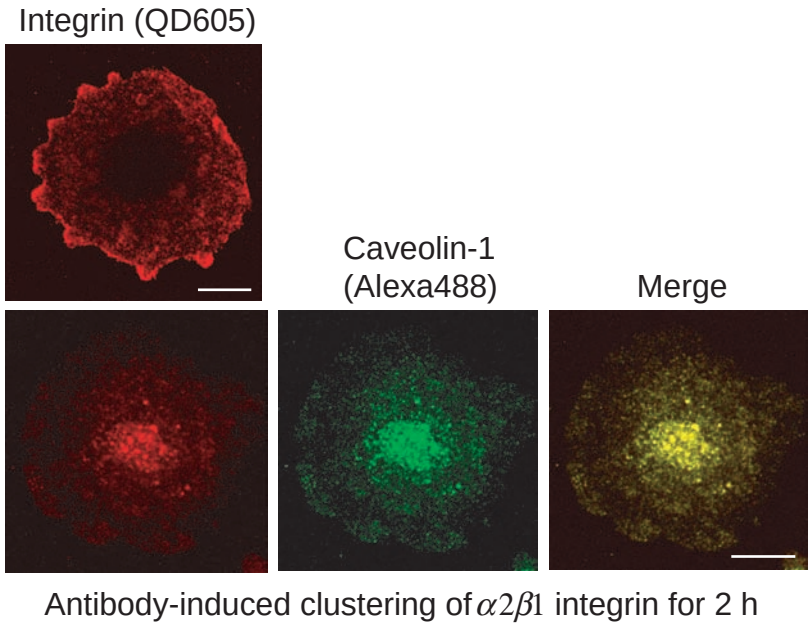


Fig. 2. $\alpha 2\beta 1$ integrin distribution in SAOS $\alpha 2\beta 1$ cells under normal conditions. (**A**) and after antibody-induced cross-linking (**B**). Integrin is labeled by monoclonal antibodies followed by goat anti-mouse quantum dots 605 (Chemicon). $\alpha 2\beta 1$ integrin is found on the plasma membrane in normal conditions (**A**). For antibody-induced cross-linking (**B**), cells are treated here with quantum dot secondary antibodies for 2 h at 37°C. After fixation, caveolin-1 is also labeled in the cells and revealed by goat anti-rabbit Alexa-488 (green). The confocal image shows that during antibody-induced cross-linking integrin accumulates very efficiently in caveolin-1 positive caveosomes in the cytoplasm. In these images the orange color of quantum dot 605 is changed into false color red in order to reveal the colocalization with caveolin-1 better. Bar = 10 μm.

that have no αI domains, for example the fibronectin receptors, bind to their ligands with a special site partially formed by their βI domains.[21] The βI domain connects the integrin α and β subunits together by interacting with a β-propeller domain in the α subunit. Structurally and functionally integrin αI and βI domains are very similar. A ligand-dependent conformational change in the βI domain triggers the leg parts of the α

and β subunits to move apart from each other, leading to the separation of the corresponding intracellular domains.[20] After the change in integrin conformation, intracellular signaling proteins can bind to cytoplasmic tails and be activated.

In the integrins that have ligand-binding αI domains, a specific glutamate residue in the α7 helix of the αI domain may act as an intrinsic ligand for the βI domain.[21] In these integrins, the ligand at first induces a change in the αI domain, followed by a similar change in the βI domain.

Integrins are connected to many different signaling pathways, which regulate cell growth, differentiation and survival. Therefore it is logical to presume that also viruses binding to integrins can activate cellular signaling. In our own studies with EV1 we noticed a rapid phosphorylation of one of the mitogen-activated protein kinases (MAP kinases), namely extracellular signal regulated kinase (ERK), already after 5 to 30 min p.i. (Upla *et al.*, 2004). MAP kinases are activated, together with induction of immediate early genes, also during the later stages of the replication cycle.[22,23] When we searched for the upstream signaling proteins it became evident that members of the PKC family activated ERK after EV1 binding, rather than the Ras pathway. Inhibition of ERK activity had no effect on the EV1 entry, but specific inhibitors of PKC-α were able to block the internalization of the virus and its receptor. PKC-α was also moderately activated during the internalization and the inhibition of its function totally blocked EV1 internalization and infection.[24]

Internalized EV1 is Found in Caveosomes

Virus entry into the host cell may occur via at least five different pathways (for a review see Refs. 25 and 26). Clathrin-dependent endocytic pathway is the best characterized internalization route and it is often referred to as the classical endocytic pathway. It mediates the constitutive entry of clathrin-coated vesicles to cytoplasm, which after rapid disintegration of the clathrin coat fuse with the peripheral early endosomes. The receptors that are attracted to enter through a clathrin-dependent pathway usually contain an internalization signal in their C-terminus, which then binds to the adaptors in the clathrin coat.

Flask-shaped caveolae have been shown to pinch off from the plasma membrane in a dynamin-dependent manner.[27,28] Caveolar endocytosis does not occur constitutively, but is rather triggered by the cargo, e.g. simian virus 40 (SV40; Ref. 29). Caveolae structures have been shown to accumulate in the cytoplasm and fuse with pre-existing caveosomes. The best markers for the caveolae are their major structural proteins, caveolin-1 and caveolin-2, which make a rope-like coating on the cytoplasmic side of these structures. Caveolin-1 is the best molecular marker for the caveosomes, too. The other characteristics of the caveosomes are the cytoplasmic location and the absence of classical endosomal markers. The morphological characteristics of caveosomes are based on cellular structures where SV40 is internalized, and it is possible that there are also other types of caveosomes.

The presence of clathrin- and caveolin-independent endocytosis pathways has been suggested on the basis of studies using cells lacking morphologically typical caveolae structures or studies in which the clathrin and caveolae pathways have been inhibited.[25,30] Similarly, some of these entry pathways are dependent on dynamin, but dynamin-independent pathways may also exist.[31]

Recently, increasing attention has been paid to the macropinocytosis pathway.[30] It is usually elicited from the raft domains on the plasma membrane. It involves ruffling of plasma membrane, which may take place rapidly after the binding of the ligand to its receptor. This ruffling seems to be dependent on Rho GTPases. Actin drives the formation of membrane ruffles, which collapse on the plasma membrane and are engulfed into the cell with a relatively large volume of liquid. It was recently suggested that carboxy-terminal binding protein 3/brefeldin A-ribosylated substrate (CtBP3/BARS) regulates macropinocytic uptake in a similar manner as dynamin regulates clathrin- and caveolae-dependent pathways.[31] CtBP3/BARS was earlier shown to mediate the fission of tubules on the Golgi.[32]

Already our first confocal microscopy studies indicated that EV1 and $\alpha 2\beta 1$ integrin do not enter the host cells via the classical endosomal structures. We found only marginal colocalization of EV1 with transferrin and no apparent colocalization with EEA1, CI-MPR or CD63, markers of early and late endosomes/lysosomes.[10] In the same

experiments it was obvious that EV1 and $\alpha 2\beta 1$ integrin are colocalized with caveolin-1 in the cytoplasm. As observed via immunoelectron microscopy, both the virus and the receptor accumulated into relatively large vesicular structures, suggesting that they could represent caveosomes. Some of the structures closely resembled the SV40-positive bundles of caveolae-type caveosomes.[29] However, the vast majority of the structures containing EV1 and $\alpha 2\beta 1$ integrin were usually rather large (300–400 nm in diameter) and showed complex morphology with tubular and vesicular contents confined by a limiting membrane (Ref. 9; Karjalainen *et al.*, unpublished). Tracing the EV1-positive structures in the electron microscopy was difficult due to the small size of the virus and the low number of viruses found in thin sections.[10] Our first hypothesis was that EV1 follows the same pathway as SV40 to the caveolae on the plasma membrane and later to the caveosomes.

There are facts proposing that in addition to caveolae, EV1 may have other more important entry mechanisms. The colocalization of EV1 and caveolin-1 is quite weak during the first minutes of entry (Ref. 33; Karjalainen *et al.*, unpublished). This has also been verified in the live cell microscopy with caveolin-1-GFP transfected cells.[33] The colocalization becomes gradually more pronounced, starting after 15 min p.i. and increasing until 2 h p.i. The internalization of EV1 is much faster than the entry of SV40, since it reaches the caveosomes 90 min earlier than SV40. When the depolymerization of microtubules is blocked and consequently the departure of SV40 from caveosomes is inhibited, the colocalization of EV1 and SV40 in caveosomes is much more pronounced. Although EV1 and SV40 can enter the same caveosome structures, the entry route may still be quite different. SV40 is transported from caveosomes to endoplasmic reticulum (ER), but there is no evidence of further traffic for EV1, suggesting that caveosomes may represent the destination prior to uncoating and release of the genome into the cytoplasm.[33]

Open Questions

Despite the fact that $\alpha 2\beta 1$ integrin has been confirmed to be a major cellular receptor for EV1 both in cell culture[34] and in transgenic

mice,[35] many open questions about the receptor function still remain. Due to the different interaction mechanism of $\alpha 2\beta 1$ integrin and EV1 when compared to natural ligands, it is not clear whether virus binding can trigger a similar conformational change in the $\alpha 2$I domain. If not, there may be neither conformational changes in the β subunit nor separation of the intracellular domains. Thus, after virus binding, the integrin signaling might be solely due to receptor clustering. A detailed structure-based analysis of $\alpha 2\beta 1$ function after EV1 binding would clarify these events.

We have shown that EV1 can be found on cell surfaces in structures resembling caveolae and later inside the cells in caveosomes.[10] However, it is not clear how many different entry routes EV1 can utilize. Furthermore, the exact timing and the mechanisms of viral uncoating and the release of the RNA genome are currently unknown. We have preliminary evidence that EV1 may also be engulfed in large macropinosome-like vesicular structures that very quickly fuse with the pre-existing caveosomes (Karjalainen *et al.*, unpublished). This would at least partly explain the different timing and routing of EV1 to the caveosomes when compared to SV40. The fluorescence *in situ* hybridization assay showed that viral RNA is present together with the capsid proteins in the caveosomes almost until the initiation of viral replication. Further studies on the molecular details of the EV1 entry process will most probably shed new light on the complexity of cellular entry pathways used by microbial pathogens and physiological ligands.

References

1. Hynes RO. (2002) Integrins: bidirectional, allosteric signaling machines. *Cell* **110**(6): 673–687.
2. Joki-Korpela P, Marjomäki V, Krogerus C, *et al.* (2001) Entry of human parechovirus 1. *J Virol* **75**(4): 1958–1967.
3. Roivainen M, Piirainen L, Hovi T, *et al.* (1994) Entry of coxsackievirus A9 into host cells: specific interactions with alpha v beta 3 integrin, the vitronectin receptor. *Virology* **203**(2): 357–365.
4. Verdaguer N, Mateu MG, Andreu D, *et al.* (1995) Structure of the major antigenic loop of foot-and-mouth disease virus complexed with a

neutralizing antibody: direct involvement of the Arg-Gly-Asp motif in the interaction. *Embo J* **14**(8): 1690–1696.
5. Wickham TJ, Mathias P, Cheresh DA, Nemerow GR. (1993) Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell* **73**(2): 309–319.
6. Williams CH, Kajander T, Hyypiä T, *et al.* (2004) Integrin alpha v beta 6 is an RGD-dependent receptor for coxsackievirus A9. *J Virol* **78**(13): 6967–6973.
7. Varga MJ, Weibull C, Everitt E. (1991) Infectious entry pathway of adenovirus type 2. *J Virol* **65**(11): 6061–6070.
8. White DJ, Puranen S, Johnson MS, Heino J. (2004) The collagen receptor subfamily of the integrins. *Int J Biochem Cell Biol.* **36**: 1405–1410.
9. Bergelson JM, Shepley MP, Chan BM, *et al.* (1992) Identification of the integrin VLA-2 as a receptor for echovirus 1. *Science* **255**(5052): 1718–1720.
10. Marjomäki V, Pietiäinen V, Matilainen H, *et al.* (2002) Internalization of echovirus 1 in caveolae. *J Virol* **76**(4): 1856–1865.
11. Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC. (2000) Structural basis of collagen recognition by integrin alpha2beta1. *Cell* **101**(1):47–56.
12. Elices MJ, Hemler ME. (1989) The human integrin VLA-2 is a collagen receptor on some cells and a collagen/laminin receptor on others. *Proc Natl Acad Sci USA* **86**(24): 9906–9910.
13. Knight CG, Morton LF, Onley DJ, *et al.* (1998) Identification in collagen type I of an integrin alpha2 beta1-binding site containing an essential GER sequence. *J Biol Chem* **273**(50): 33287–33294.
14. Bergelson JM, Chan BM, Finberg RW, Hemler ME. (1993) The integrin VLA-2 binds echovirus 1 and extracellular matrix ligands by different mechanisms. *J Clin Invest* **92**(1): 232–239.
15. King SL, Kamata T, Cunningham JA, *et al.* (1997) Echovirus 1 interaction with the human very late antigen-2 (integrin alpha2beta1) I domain. Identification of two independent virus contact sites distinct from the metal ion-dependent adhesion site. *J Biol Chem* **272**(45): 28518–28522.
16. Xing L, Huhtala M, Pietiäinen V, *et al.* (2004) Structural and functional analysis of integrin alpha2I domain interaction with echovirus 1. *J Biol Chem* **279**(12): 11632–11638.

17. Filman DJ, Wien MW, Cunningham JA, *et al.* (1998) Structure determination of echovirus 1. *Acta Crystallogr D Biol Crystallogr* **54**: 1261–1272.
18. Emsley J, King SL, Bergelson JM, Liddington RC. (1997) Crystal structure of the I domain from integrin alpha2beta1. *J Biol Chem* **272**(45): 28512–28517.
19. Xiong JP, Stehle T, Diefenbach B, *et al.* (2001) Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science* 294(5541): 339–345.
20. Xiao T, Takagi J, Coller BS, *et al.* (2004) Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. *Nature* **432**(7013): 59–67.
21. Alonso JL, Essafi M, Xiong JP, *et al.* (2002) Does the integrin alphaA domain act as a ligand for its betaA domain? *Curr Biol* **12**(10): R340–342.
22. Huttunen P, Heino J, Hyypiä T. (1997) Echovirus 1 replication, not only virus binding to its receptor, VLA-2, is required for the induction of cellular immediate-early genes. *J Virol* **71**(5): 4176–4180.
23. Huttunen P, Hyypiä T, Vihinen P, *et al.* (1998) Echovirus 1 infection induces both stress- and growth-activated mitogen-activated protein kinase pathways and regulates the transcription of cellular immediate-early genes. *Virology* **250**(1): 85–93.
24. Upla P, Marjomäki V, Kankaanpää P, *et al.* (2004) Clustering induces a lateral redistribution of alpha 2 beta 1 integrin from membrane rafts to caveolae and subsequent protein kinase C-dependent internalization. *Mol Biol Cell* **15**(2): 625–636.
25. Damm EM, Pelkmans L, Kartenbeck J, *et al.* (2005) Clathrin- and caveolin-1-independent endocytosis: entry of simian virus 40 into cells devoid of caveolae. *J Cell Biol* **168**(3): 477–488.
26. Smith AE, Helenius A. (2004) How viruses enter animal cells. *Science* **304**(5668): 237–242.
27. Henley JR, Krueger EW, Oswald BJ, McNiven MA. (1998) Dynamin-mediated internalization of caveolae. *J Cell Biol* **141**(1): 85–99.
28. Oh P, McIntosh DP, Schnitzer JE. (1998) Dynamin at the neck of caveolae mediates their budding to form transport vesicles by GTP-driven fission from the plasma membrane of endothelium. *J Cell Biol* **141**(1): 101–114.
29. Pelkmans L, Kartenbeck J, Helenius A. (2001) Caveolar endocytosis of simian virus 40 reveals a new two-step vesicular-transport pathway to the ER. *Nat Cell Biol* **3**(5): 473–483.

30. Kirkham M, Parton RG. (2005) Clathrin-independent endocytosis: New insights into caveolae and non-caveolar lipid raft carriers. *Biochim Biophys Acta*. **1745** (3): 273–86.
31. Bonazzi M, Spano S, Turacchio G, *et al.* (2005) CtBP3/BARS drives membrane fission in dynamin-independent transport pathways. *Nat Cell Biol* 7(6): 570–580.
32. Weigert R, Silletta MG, Spano S, *et al.* (1999) CtBP/BARS induces fission of Golgi membranes by acylating lysophosphatidic acid. *Nature* **402**(6760): 429–433.
33. Pietiainen V, Marjomäki V, Upla P, *et al.* (2004) Echovirus 1 endocytosis into caveosomes requires lipid rafts, dynamin II, and signaling events. *Mol Biol Cell* **15**(11): 4911–4925.
34. Bergelson JM, Shepley MP, Chan BM, *et al.* (1992) Idetification of the intergin VLA-2 as a receptor for echovirus 1. *Science* **255**(5052): 1718–1720.
35. Hughes SA, Thaker HM, Racaniello VR. (2003) Transgenic mouse model for echovirus myocarditis and paralysis. *Proc Natl Acad Sci USA* **100**(26): 15906–15911.
36. Smith AE, Helenius A. (2004) How viruses enter animal cells. *Science* **304**(5668): 237–242.

This page intentionally left blank

Chapter 4

Entry Mechanism of Murine and SARS Coronaviruses; Similarity and Dissimilarity

*Fumihiro Taguchi**

Introduction

The Coronaviruses consist of human and animal viruses that infect a variety of animal species, as well as different organs in the hosts.[1,2] Since the causative agent of a life-threatening pneumonia, severe acute respiratory syndrome (SARS), was revealed to be a member of the coronavirus group,[3,4] much attention has been focused on these viruses, from both the academic and non-academic societies. SARS is a new type of infectious pneumonia that emerged in the Gangdong province of China in November 2002.[5] It spread around the world through an outbreak in Hong Kong in March 2003. Patients were reported in 29 countries, and around 800 people died, although a majority of these cases were in Asian countries. Within a month after the virus that causes SARS was isolated, its entire genome sequence became available, which is attributable to the collaboration between many institutes around the world.[6,7] The result is the acknowledgement of a novel coronavirus, now called SARS coronavirus (SARS-CoV), that is different from other already known existing human and animal coronaviruses.[3,4,6,7] Since a number of scientists from distinct fields have participated in the research on this emerging virus,

*Laboratory of Respiratory Viral Diseases and SARS, National Institute of Infectious Diseases, Murayama Branch, 4-7-1 Gakuen, Musashi-Murayama, Tokyo 208-0011, Japan.

a plenitude of information has become available, and now SARS-CoV has become one of the best-studied members among the coronaviruses.[8]

SARS-CoV was initially reported to be different from any known coronaviruses and thus failed to be classified in any of the coronavirus groups.[4] However, a detailed phylogenetical study including Torovirus that is also classified in *Coronaviridae*, suggests it to be a member of group 2, though it is most distant from the other members of this group.[9] It is also reported that the SARS-CoV genome is a mosaic of mammalian and avian coronaviruses.[10] Most coronaviruses, other than SARS-CoV, do not cause a fatal disease. Two species of human coronavirus are causative agents of the common cold, and studies on these viruses were restricted because of their weak pathogenic nature.[1,2] Other coronaviruses infect domestic animals, pets and experimental animals, but mostly do not cause severe and fatal diseases. However, as they are important pathogens from an economic point of view, these animal coronaviruses have been researched extensively.[2]

A prototype coronavirus, mouse hepatitis virus (MHV), has been studied as a model of human disease, such as hepatitis and demyelinating encephalomyelitis.[2] This virus is one of the most studied coronaviruses, especially in terms of its molecular mechanism of viral replication as well as its cell entry.[2,11] In this review, I would like to describe how MHV and SARS-CoV interact with their respective receptors and enter cells, as well as the similarity and dissimilarity between these two coronaviruses.

S Proteins of Coronaviruses

Coronaviruses have morphologically characteristic spikes resembling a solar corona on the virion surface, for which they were so named. The spike consists of a globular structure in the outer extremity and a stem-like structure beneath it. The spike is thought to be composed of three molecules of spike (S) protein. The S protein is a class I fusion protein with a molecular weight of 180 to 200 kDa (Fig. 1).[2] After being synthesized, MHV and some other coronavirus S proteins are cleaved around the middle of the molecule into two subunits, N-terminal S1 and C-terminal S2, by host cell-derived proteases.[12] There is a high similarity in the amino acid sequences in S2, but these are quite variable in S1

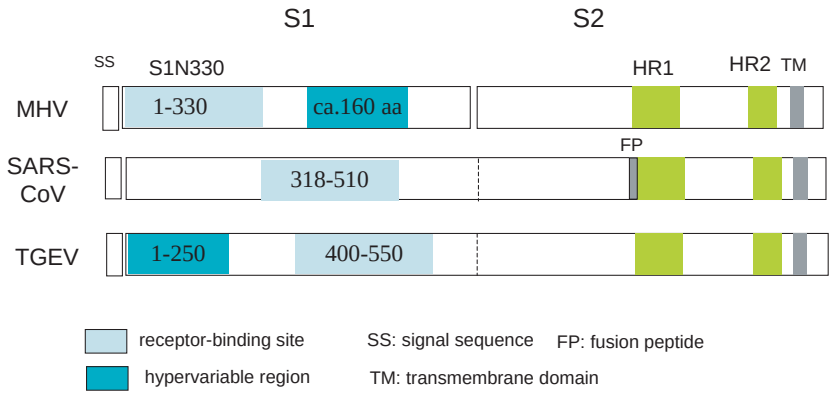


Fig. 1. Structural comparison of different coronavirus S proteins. Coronavirus S protein is a class I fusion protein, which is similar to the influenza HA and HIV gp 160 envelope protein. The MHV S protein is cleaved in the middle of the molecule, resulting in N terminal S1 and C terminal S2 subunits, while the SARS-CoV S protein is not cleaved. The N terminal region of S1, consisting of 330 residues (S1N330), is a receptor-binding site for MHV, whereas the internal region from residues 318 to 510 is responsible for receptor binding of SARS-CoV S protein. Downstream from the S1N330 exists a hypervariable region consisting of ~160 residues, whereas various deletions exist in N terminal region consisting of ca. 250 residues of porcine transmissible gastroenteritis virus (TGEV) S protein. In S2, there are two heptad repeats, HR1 and HR2, that play critical roles in viral envelope-cell membrane fusion. Fusion peptide is located close upstream of HR1 in SARS-CoV S protein; however, the MHV S protein fusion peptide is located elsewhere in an internal, as yet unidentified, region of S2. The S protein of SARS-CoV is more similar to the TGEV S protein than to MHV S in terms of the cleavability and localization of the receptor-binding site.

when comparison is made among the various MHV strains. The cleavage site consists of a basic amino acid cluster, which is recognized by cellular proteases such as furin. The S1 fragment is believed to construct the globular part of the spike and the S2 fragment its stem-like portion.[13] SARS-CoV S protein, however, is not cleaved (Fig. 1).[6,7] This feature of uncleavability is similar to that of the S protein of the coronaviruses classified into group 1, such as porcine transmissible gastroenteritis virus (TGEV) and feline infectious peritonitis virus (FIPV).[2] In the middle of the SARS-CoV S protein, there are regions containing two basic amino

acid residues, resembling the cleavage site found in the MHV S protein.[6,7] However, this region is not utilized for cleavage events in the de novo-synthesized S protein, probably due to an unsatisfactory cleavage motif consisting of a small number of basic amino acids. Cleavage of the S protein is not always necessary for fusion activity of the MHV strains[14]; however, it is important for a highly virulent strain of MHV, MHV-2.[15]

The N terminal region, consisting of 330 amino acid residues in the MHVS S1 subunit, is a receptor-binding site.[16,17] Downstream is a hypervariable region where mutations and deletions of various numbers of amino acids arise during repeated passages of the virus.[18] In contrast, SARS-CoV S protein has its receptor-binding domain not at the N terminal, but rather in an internal region consisting of 193 residues at position 318 to 510 (Fig. 1).[19] This localization of the receptor-binding site is similar to that of group 1 coronavirus S protein, e.g., that of TGEV or human coronavirus 229E; the S proteins of these viruses have the receptor-binding domain at the position of 400–550 amino acid from the initiator of the S protein.[20] Thus, the SARS-CoV S protein is more similar to group 1 coronavirus S protein than to MHV S protein, in terms of the cleavability and localization of the receptor-binding site. In contrast, there is a high similarity in the S2 subunit of MHV and the corresponding region in SARS-CoV S protein. They are also very similar structurally and functionally to the membrane-anchored subunit of HIV and influenza, gp41 and HA2.[21] All of these proteins contain two heptad repeats (HR) in the upstream of the transmembrane domain.[22–24] The HR located upstream is named HR1 and that located downstream is called HR2. These heptad repeats play an important role in virus-cell membrane fusion, namely, viral entry into cells. A fusion peptide, a hydrophobic amino acid cluster of coronavirus that penetrates into cell membrane at an initial phase of virus-cell membrane fusion, is not located in the N terminal region but rather in the internal not-yet identified region of the membrane-anchored subunit of MHV S2, whereas the SARS-CoV fusion peptide is reported in close upstream of HR1.[22–24]

Coronavirus Receptors

There are at least three different types of molecule reported to work as receptors for the various coronaviruses. The receptor protein first

identified is a CEACAM1 (carcinoembryonic antigen cell adhesion molecule 1) that serves as an MHV receptor. CEACAM1 is classified in the immunoglobulin (Ig) superfamily.[25,26] The prototype CEACAM1 is composed of four Ig-like ectodomains (in the order of N, A1, B and A2 from the N terminus), a transmembrane domain, and a cytoplasmic tail (Cy). There are four isoforms, two of which have 4 Ig-like domains, and the other two, two Ig-like domains (N and A2), one of which has either a long or short Cy.[27] Allelic forms are known for CEACAM1, CEACAM1^{a} and CEACAM1^{b}, the former being expressed in most laboratory mouse strains susceptible to MHV and the latter in SJL relatively resistant to MHV.[27,28] CEACAM1 is expressed on various epithelial cells, endothelial cells and hemopoietic cells.[29] It functions as a cell adhesion molecule,[30] a signaling molecule and an angiogenic factor.[31]

The virus-binding site of CEACAM1 is located in the N domain.[32,33] Crystal structural analysis of CEACAM1 revealed that a uniquely folded CC′-loop that protrudes from the N domain mass plays an important role in virus binding.[34] This region also participates in homophilic adhesion activity that is retained by CEACAM1.[34]

The receptor for SARS-CoV was identified as an angiotensin-converting enzyme 2 (ACE2).[35] ACE2 is a type I integral membrane protein with carboxypeptidase activity and plays an important role in the rennin angiotensin system for blood pressure homeostasis. ACE2 is expressed in a variety of organs, such as the lungs and intestines, the main target organs of SARS-CoV, as well as in the heart or kidneys where SARS-CoV infection is rarely observed or limited.[36] As for the virus binding site on ACE2, it was postulated from the homology to other proteins (whose crystal structure was studied), that a negatively charged ridge close to the deep channel containing a catalytic site is the virus-binding site.[37] Recently, the more precise binding site was determined to locate to α-helix 1 of ACE2 and to a loop leading to β-sheets.[38]

The other coronavirus receptor is aminopeptidase N (APN) that serves as a receptor for group I coronaviruses.[39] Being similar to ACE2, APN is a zinc metalloprotease. Group 1 coronaviruses infecting different species of animals utilize APN of the host species, though feline APN serves as a receptor for porcine, human and canine coronaviruses as well.[39]

Cell Entry Pathways of Coronaviruses

Syncytium formation is generally observed in cultured cells infected with various MHV strains under a wide range of pH environments. This is attributed to the S protein, as cells expressing the MHV S protein alone are fused, similar to their infections.[40] The replication of MHV is, in general, hardly affected by lysosomotropic agents.[41] These observations suggest that MHV enters the cells from the cell surface after binding to its receptor. Thus, MHV is thought to take a non-endosomal pathway (Fig. 2). Exceptions are also reported. A mutant virus isolated from persistent infection fails to induce fusion formation under neurtral pH, but it causes cell fusion when the cells are treated by low pH.[41] This situation is very similar to that found in cases of influenza or VSV fusion formation,[21] and thus suggested that this particular virus takes an endosomal pathway; in the endosomal acidic environment, the S protein is fusogenically activated. MHV-2, a non-syncytium formerly under a wide range of pH culture conditions, has a unique feature for fusion formation. Infected cells undergo fusion formation when the cells are treated with trypsin,[42] which is very similar to the case with SARS-CoV fusion formation.[43,44]

The SARS-CoV entry pathway is thought to be via endosomes — i.e., the endosomal pathway[43] — though the entry mechanism of this virus seems to be different from that of influenza or VSV (Fig. 2). SARS-CoV infection in cultured cells does not induce syncytium formation under a wide range of pH environments, even under low pH conditions, but cell fusion can be induced when the infected cells are treated with trypsin.[43,44] The environmental pH does not affect the fusion activity of trypsin. Moreover, SARS-CoV infection is affected by a lysosomtropic agent, such as bafilomycin or ammonium chloride.[43] These findings suggest that SARS-CoV is trafficked to the endosome after binding to ACE2, where the S protein is fusogenically activated by the trypsin-like protease which is presumably supplied by the lysosome fused with endosome. Recently, cathepsin L was reported to be a possible protease to activate the SARS-CoV S protein for fusion.[45] Trypsin treatment produces a fragment of about 100 kDa of S from ca. 180-kDa uncleaved S protein. This 100-kDa fragment corresponds to S2. The transition from 180 kDa to 100 kDa of S protein is caused by not only trypsin, but also by other proteases such as thermolysin or

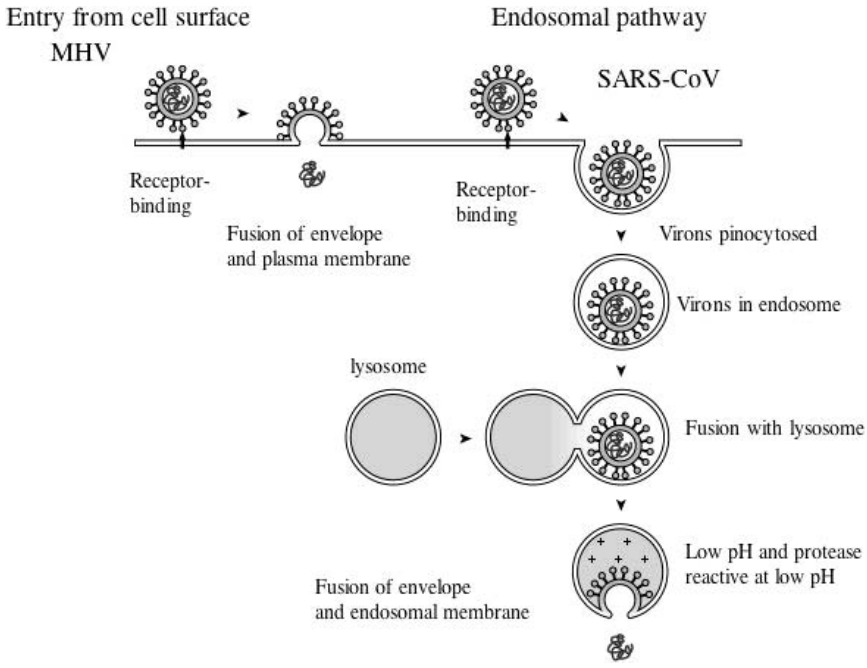


Fig. 2. Entry pathway of MHV and SARS-CoV. MHV S protein binds its receptor CEACAM1 on the cell surface, and fusion of viral envelope and plasma membrane takes place. Then, the genome is internalized directly into the cytoplasm through a plasma membrane. Possible entry mechanism of SARS-CoV is as follows. SARS-CoV S protein binds ACE2, and the virion is incorporated into the endosome by pinocytosis/endocytosis. Then, the endosome is fused with lysosome, which provides the endosome with various proteases. Protease, cathepsin L, reactive at a low pH environment activates the fusogenicity of SARS-CoV S protein, facilitating viral envelope and endosomal membrane fusion. Then, the SARS-CoV genome enters the cells via an endosomal membrane.

elastase. These proteases also induce fusion of SARS-CoV-infected cells. The S proteins treated with these proteases show an indistinct size of S protein, ca. 100 kDa.[44] All of these data suggest that a protease reactive under low pH environment, most likely cathepsin L, plays an important role in the entry mechanism of SARS-CoV.

Recently, we found that proteases that induce S protein cleavage as well as fusion activity facilitate the entry of SARS-CoV from cell surface by affecting the virion S proteins that bound to the receptor.[44] This

finding is in good agreement with the hypothesis proposed by Simmons and his colleagues that SARS-CoV takes an endosomal pathway and enters cells after cleavage of the S protein with protease in the endosome.[43,45] Interestingly, entry from cell surface mediated by proteases results in more efficient infection than the infection via the endosomal pathway.[44]

Mechanism of Cell Entry

The cell entry mechanisms of MHV and SARS-CoV are thought to be similar to that of a well-known mechanism found in HIV or influenza virus studies (Fig. 3).[46] A portion of S1N330 of the MHV S protein binds to the CC´-loop in the N domain of CEACAM1, which removes the S1 fragment from the envelope-anchored S2.[47,48] A fusion peptide in S2 then penetrates the plasma membrane of the target cells, which would trigger the conformational changes of S2.[49] The conformational changes involve the formation of the so-called six helix bundles, composed of a trimeric coiled-coil inside HR1 and three copies of HR2 that overhang onto the coiled-coil inside HR1s. These conformational changes of S2 place the viral envelope and cell membrane in close proximity and facilitate the fusion of these two membranes.

SARS-CoV S protein binds to ACE2 with an internal region of an S1 counterpart. After transport to the endosome, the S protein is presumably cleaved by a protease, most likely cathepsin L which is functional in a low pH environment in the endosome. The conformational changes similar to that of the MHV S protein could follow. The fusion of the envelope and endosomal membrane then takes place and the genome of SARS-CoV is internalized into the cell cytoplasm. Six helix bundles are formed with the SARS-CoV S protein and a peptide corresponding to HR2 blocks its infection, as was shown in HIV and MHV infections.[23,46] Thus, the fusion mechanism of the SARS-CoV envelope and cellular membrane is thought not to differ from those reported thus far in other enveloped viruses; however, the fusogenic activation process of the SARS-CoV S protein is quite unique in that it requires a protease for this activation.

Receptor-independent Entry Mechanism

A particular murine coronavirus, MHV-JHMV, is believed to enter cells in a receptor-independent fashion. This observation was first reported by

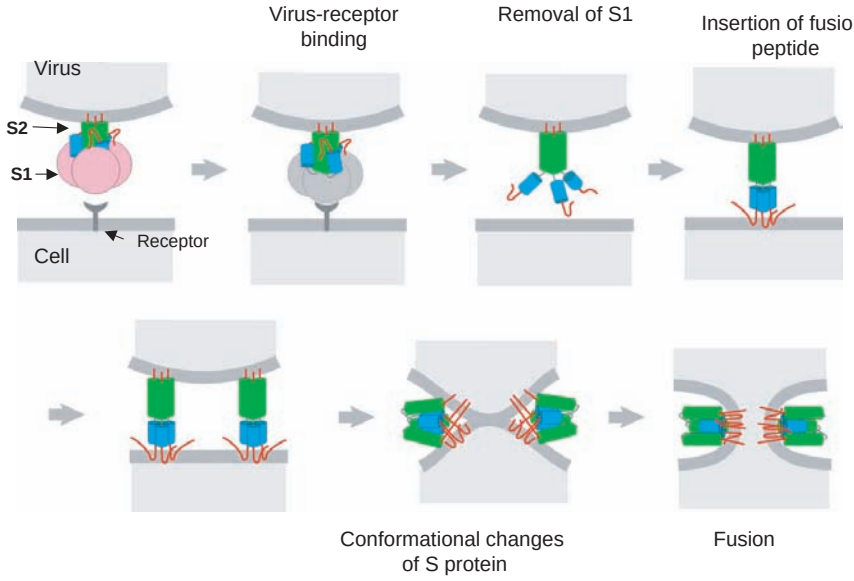


Fig. 3. Entry mechanism of coronaviruses. The entry mechanism of MHV is very similar to that of HIV. MHV S protein binds its receptor CEACAM1, and then S1 is removed from the envelope-anchored S2 subunit. Fusion peptide is inserted into the plasma membrane. These events trigger conformational changes of the S2 to form 6 helix bundles consisting of trimeric coiled coil HR1, over which 3 copies of HR2 bind. The fusogenic activation of SARS-CoV S protein is supposedly conducted by cellular proteases, most likely cathepsin L, although a detailed mechanism of this activation is not yet understood. Viral and cell membrane fusion is carried out in a similar fashion to that of HIV and MHV, since 6 helix bundles were shown to be formed.

Gallagher and his colleagues as a spread of JHMV infection — from cells initially infected in a receptor-mediated fashion — to those cells lacking the MHV receptor CEACAM1.[50] They showed this phenomenon by overlaying JHMV-infected cells onto a CEACAM1 negative cell monolayer. Its S protein is responsible for this infection.[50] The mechanism of receptor-independent infection is thought to be due to a unique feature of the JHMV S protein. As described above, the S1 subunit of MHV is removed from S2 after binding its receptor; S1 and S2 are not covalently linked and thus S1 is physically removed from S2 under certain

conditions, such as those following ultracentrifugation. The S1 subunit of JHMV is particularly removable, which takes place as a naturally occurring event.[51,52] This characteristic is thought to be responsible for a receptor-independent fusion mechanism. The S2 subunit, from which S1 is removed in a receptor-unmediated fashion, could undergo conformational changes and is fusogenically activated in a way similar to that found when S1 binds to its receptor and becomes detached from S2.[49] A critical condition for the receptor-independent fusion/infection is that this activation process of the S protein should be carried out in close proximity of the target cell membrane. Most recently, this hypothesis was strengthened by the infection of spinoculation; spinning the inoculated viruses together with the target receptor-negative cells resulted in the viruses being attached on the cell surface.[52] The receptor-independent JHMV infection was suggested to result in its high neurovirulence in mice.[53] Although it is not so remarkable, some strains of MHV other than JHMV retain the ability of receptor-independent infection.[54]

Concluding Remarks

Although different types of protein are used as receptors for MHV or SARS-CoV, the cell entry mechanism of these viruses looks very similar to that first postulated as an entry mechanism for the influenza virus and HIV. However, the fusogenic activation of the SARS-CoV S protein by the cell-derived protease under low pH environment is a quite unique phenomenon. MHV-2 could share this feature and provide an excellent model for understanding the SARS-CoV entry mechanism. Detailed molecular studies on the S protein and receptor interactions, as well as the subsequent cell entry mechanism, will facilitate the development of anti-SARS-CoV reagents.

Acknowledgments

I greatly appreciate Shutoku Matsuyama for his valuable and stimulating discussions and for illustrations used in this review. I also thank Keiko Nakagaki, Miyuki Kawase as well as Shigeru Morikawa and his group members for their continuous discussions and encouragement. This work is financially supported by grants from the Ministry of Education, Culture, Sports, Science and Technology (16017308),

from the Ministry of Health, Labor and Welfare (H16-Shinkoh-9) and from the Human Science Foundation (KH51050).

References

1. Tyrell DA, Almeida JD, Berry DM, *et al.* (1968) Coronaviruses. *Nature* (Lond.) **220**: 650.
2. Lai MMC, Cavanagh D. (1997) The molecular biology of coronaviruses. *Adv Virus Res* **48**: 1–100.
3. Drosten C, Gunter S, Priser W, *et al.* (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *New Eng J Med* **348**: 1967–1976.
4. Ksiazek TG, Erdman D, Goldsmith D, *et al.* (2003) A novel coronavirus associated with severe acute respiratory syndrome in Canada. *New Engl J Med* **348**: 1953–1966.
5. The Chinese SARS Molecular Epidemiology Consortium. (2004) Molecular Evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* **303**: 1666–1669.
6. Marra MA, Jones SJM, Astell CR, *et al.* (2003) The genome sequence of the SARS-associated coronavirus. *Science* **300**: 1399–1404.
7. Rota PA, Oberste MS, Monroe SS, *et al.* (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**: 1394–1399.
8. Peiris JS, Guan Y, Yuen KY. (2005) Severe acute respiratory syndrome. *Nat Med* **10**: 588–597.
9. Snijder EJ, Bredenbeek PJ, Dobbe JC, *et al.* (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* **331**: 991–1004.
10. Stavrinides J, Guttman DS. (2004) Mosaic evolution of the sevee acute respiratory syndrome coronavirus. *J Virol* **78**: 76–82.
11. Gallgher T, Buchmeier M. (2001) Coronavirus spike proteins in viral entry and pathogenesis. *Virology* **279**: 371–374.
12. Sturman LS, Ricard CA, Holmes KV. (1985) Proteolytic cleavage of the E2 glycoprotein of murine coronavirus: activation of cell-fusing activity of virions by trypsin and separation of two different 90K cleavage fragments. *J Virol* **56**: 904–911.
13. De Groot RJ W, Luytjes MC, Horzinek BAM, *et al.* (1987) Lenstra. Evidence for a coiled-coil structure in the spike of coronaviruses. *J Mol Biol* **196**: 963–966.
14. Taguchi F. (1993) Fusion formation by uncleaved spike protein of murine coronavirus JHMV variant cl-2. *J Virol* **67**: 1195–1202.

15. Yamada YK, Takimoto K, Yabe M, Taguchi, F. (1997) Acquired fusion activity of a murine coronavirus MHV-2 variant with mutations in the proteolytic cleavage site and the signal sequence of the S protein. *Virology* **227**: 215–219.
16. Kubo H, Yamada YK, Taguchi F. (1994) Localization of neutralizing epitopes and the receptor-binding site within the amino-ternimal 330 amino acids of the murine coronavirus spike protein. *J Virol* **68**: 5403–5410.
17. Suzuki H, Taguchi F. (1996) Analysis of the receptor binding site of murine coronavirus spike glycoprotein. *J Virol* **70**: 2632–2636.
18. Parker SE, Gallagher TM, Buchmeier MJ. (1989) Sequence analysis reveals extensive polymorphism and evidence of deletions within the E2 glycoprotein gene of several strains of murine hepatitis virus. *Virology* **173**: 664–673.
19. Wong SK, Li W, Moore MJ, *et al.* (2003) A 193-amino-acid fragment of the SARS coronavirus s protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem* **279**: 3197–3201.
20. Bonavia A, Zelus BD, Wentworth DE, *et al.* (2003) Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus HcoV-229E. *J Virol* 77: 2530–2538.
21. White JM. (1990). Viral and cellular membrane fusion protein. *Ann Rev Physiol* **52**: 675–695.
22. Bosch BJ, van der Zee R, de Haan CAM, *et al.* (2003) The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* 77: 8801–8811.
23. Bosch BJ, Martina BE, Van Der Zee R, *et al.* (2004) Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *Proc Natl Acad Sci USA* **101**: 8455–8460.
24. Tripet B, Howard MW, Jobling M, *et al.* (2004) Structural characterization of the SARS-coronavirus spike S fusion protein core. *J Biol Chem* **279**: 20836–20849.
25. Williams RK, Jiang GS, Holmes KV. (1991) Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proc Natl Acad Sci USA* **88**: 5533–5536.
26. Dveksler GS, Pensiero MN, Cardellichio CB, *et al.* (1991) Cloning of the mouse hepatitis virus (MHV) receptor: expression in human and hamster cell lines confers susceptibility to MHV. *J Virol* **65**: 6881–6891.
27. Beauchemin N, Draber P, Dveksler G, *et al.* (1999) Nomenclature announcement: redefined nomenclature for members of the carcinoembryonic antigen family. *Exp Cell Res* **252**: 243–249.

28. Yokomori K, Lai MMC. (1992) The receptor for mouse hepatitis virus in the resistant mouse strain SJL is functional: implications for the requirement of a second factor for viral infection. *J Virol* **66**: 6931–6938.
29. Godfraind C, Langreth SG, Cardellichio CB, *et al.* (1995) Tissue and cellular distribution of an adhesion molecule in the carcinoembryonic antigen family that serves as a receptor for mouse hepatitis virus. *Lab Investig* **73**: 615–627.
30. McCuaig K, Turbide C, Beauchemin N. (1992) MmCGM 1a: a mouse carcinoembryonic antigen gene family member, generated by alternative splicing, functions as an adhesion molecule. *Cell Growth Differ* **3**: 165–174.
31. Obrink B. (1997) CEA adhesion molecules: multifunctional proteins with signal-regulatory properties. *Curr Opin Cell Biol* **9**: 616–626.
32. Dveksler GS, Pensiero MN, Dieffenbach CW, *et al.* (1993) Mouse hepatitis virus strain A59 and blocking antireceptor monoclonal antibody bind to the N-terminal domain of cellular receptor. *Proc Natl Acad Sci USA* **90**: 1716–1720.
33. Miura HS, Nakagaki K, Taguchi F. (2003) N terminal domain of murine coronavirus receptor CEACAM1 is responsible for fusogenic activation and conformational changes of the spike protein. *J Virol* **78**: 216–223.
34. Tan K, Zelus BD, Meijers R, *et al.* (2002) Cristal structure of murine sCEACAM1a[1,4]: a coronavirus receptor in the CEA family. *EMBO J* **21**: 2076–2086.
35. Li W, Moore MJ, Vasilleva N, *et al.* (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**: 450–454.
36. Harmer D, Gilbert M, Borman R, Clark KL. (2002) Quantatitve mRNA expression profiling of ACE 2, a novel homologue of angiotensin converting enzyme. *FEBS Lett* **532**: 107–110.
37. Prabakaran P, Xiao X, Dimitrov DS. (2004) A model of the ACE2 structure and function as a SARS-CoV receptor. *Biochem Biophys Res Com* **314**: 235–241.
38. Li W, Zhang C, Sui J, *et al.* (2005) Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J* **24**: 1634–1643.
39. Tresnan DB, Levis R, Holmes KV. (1996) Feline aminopeptidase N serves as a receptor for feline, canine, porcine, and human coronaviruses in serogroup I. *J Virol* **70**: 8669–8674.
40. Taguchi F, Ikeda T, Shida H. (1992) Molecular cloning and expression of a spike protein of neurovirulent murine coronavirus JHMV variant cl-2. *J Gen Virol* **73**: 1065–1072.

41. Gallagher TM, Escarmis C, Buchmeier MJ. (1991) Alteration of the pH dependence of coronavirus-induced cell fusion: effect of mutations in the spike glycoprotein. *J Virol* **65**: 1916–1928.
42. Yoshikura H, Tejima S. (1981) Role of protease in mouse hepatitis virus-induced cell fusion. *Virology* **113**: 503–511.
43. Simmons G, Reeves JD, Rennekamp AJ, *et al.* (2004) Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proc Natl Acad Sci USA* **101**: 4240–4245.
44. Matsuyama S, Ujike M, Morikawa S, *et al.* (2005) Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection. *Proc Natl Acad Sci USA* **102**: 12543–12547.
45. Simmons G, Gosalia DN, Rennekamp AJ, *et al.* (2005) Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci USA* **102**: 11876–11881.
46. Chen D, Kim PS. (1998) HIV entry and its inhibition. *Cell* **93**: 681–684.
47. Gallagher TM. (1997) A role for naturally occurring variation of the murine coronavirus spike protein in stabilizing association with the cellular receptor. *J Virol* **71**: 3129–3137.
48. Saeki K, Ohtsuka N, Taguchi F. (1997) Identification of spike protein residues of murine coronavirus responsible for receptor-binding activity by use of soluble receptor-resistant mutants. *J Virol* **71**: 9024–9031.
49. Matsuyama S, and Taguchi F. (2002) Receptor-induced conformational changes of murine coronavirus spike protein. *J Virol* **76**: 11819–11826.
50. Gallagher T, Buchmeier MJ, Perlman S. (1992) Cell-receptor independent infection by a neurotropic murine coronavirus. *Virology* **191**: 517–522.
51. Krueger DK, Kelly SM, Lewicki DN, *et al.* (2001) Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. *J Virol* **75**: 2792–2802.
52. Watanabe R, Matsuyama S, Taguchi F. (2006) Receptor-independent infection of murine coronavirus: analysis by spinoculation. *J Virol* **80**: 4901–4908.
53. Nakagaki K, Nakagaki K, Taguchi F. (2005) Receptor-independent spread of a highly neurotropic murine coronavirus JHMV from initially infected microglial cells in mixed neural culture. *J Virol* **79**: 6102–6110.
54. Taguchi F, Matsuyama S, Saeki K. (1999) Difference in Bgp-independent fusion activity among mouse hepatitis viruses. *Arch Virol* **144**: 2041–2049.

Chapter 5

Hepatitis Viruses, Signaling Events and Modulation of the Innate Host Response

*Syed Mohammad Moin, Anindita Kar-Roy and Shahid Jameel**

Viruses and their vertebrate hosts have co-evolved for millions of years. While the host has developed an immune system to deal with invading pathogens, including viruses, the latter have developed counter strategies. The host response comprises an early nonspecific innate system and an antigen-specific adaptive immune system. Accordingly, viruses have also evolved strategies to overcome these, based largely on modulation of host signaling pathways. Here we outline components of the innate host response and cell death and survival strategies, and summarize work on how hepatitis viruses modulate these responses. At least five separate viral agents are known to cause hepatitis, leading to significant human morbidity and mortality. Their spectrum ranges from acute and self-limited infections to those that become chronic and stay with the host for its entire lifespan. This varied presentation of viral hepatitis is rationalized here in terms of the virus' ability to deal with the innate host response.

Viral hepatitis is an ancient disease that had been recorded in early Babylonian, Greek and Chinese scriptures.[1] "Hepatitis," which means inflammation of the liver, is not a single disease, but is caused by a spectrum of agents, viral or otherwise. It was not until 1969 that the first human hepatitis virus — the hepatitis B virus (HBV) — was

* Virology Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110067, India.

isolated, followed by the hepatitis A virus (HAV) in 1973. In the following years viral agents causing disease similar to HAV and HBV, but negative for tests specific for these viruses, were reported and named as non-A, non-B (NANB) hepatitis viruses, with distinct post-transfusion or enteric epidemiology.[2] The agent for post-transfusion NANB hepatitis was later identified and named as hepatitis C virus.[2] The agent responsible for ET-NANBH was first identified in 1983 as being responsible for an epidemic in New Delhi, India in 1955 in which 29,000 people fell ill; its genome was subsequently cloned and the agent was named hepatitis E virus (HEV).[3] Hepatitis D virus (HDV), or the delta agent, is a defective RNA virus that infects liver cells only if already infected by HBV. There are other viruses like cytomegalovirus, mumps, rubella and yellow fever viruses that infect the human liver, but are not considered hepatitis viruses since the primary sites of infection are extrahepatic. More than 500 million people worldwide suffer from viral hepatitis. The disease can either be self-limiting, as in the case of hepatitis A and E, or can persist, as in the case of hepatitis B and C, leading to cirrhosis and hepatocellular carcinoma (HCC).

Viruses are obligate intracellular parasites that have co-evolved with their hosts to permit virus replication without the destruction of the host cell. The host's first line of defense is through the innate response that is mediated with the help of leukocytes, monocytes, macrophages and natural killer cells. This response is nonspecific where the killer cells destroy the pathogen by phagocytosis and activation of lysosomes. In addition, induction of the inflammatory response mediated by cytokines helps in viral clearance by initiating apoptotic cell death. Adaptive immunity, on the other hand, is antigen-specific and generates immune memory. Components of adaptive immunity include antigen-specific lymphocytes such as B cells that produce neutralizing antibodies, and T cells that have helper and cytolytic effector functions. The critical balance between survival and death is dependent upon complex signaling events within a host cell. Thus the outcome of a viral infection is determined by host responses such as immunity and cell death and the ability of the virus to counteract and modulate these responses. In HAV and HEV infections, although the virus is able to establish an infection, the host's innate

and acquired antiviral immune responses are effective in clearing the virus (Mattioli 2004).[4] On the other hand, HBV and HCV can persist in the host for a long time, often the entire life of the host.[5] In such cases, there is an initial acute phase of infection when the virus employs strategies to avoid host innate defenses, escapes the acquired immunity and blocks the apoptotic clearance of infected cells to finally establish chronic infection in the host.

In this chapter we will discuss the innate defense, apoptosis and cell survival pathways and the mechanisms through which hepatitis viruses modulate these pathways to establish a successful infection in the host. An introduction to the genomic organization of these viruses and the different proteins they encode is also provided to aid and orient the reader for a better understanding of how these different proteins manipulate the host defense.

Genome Structure and Organization of Hepatitis Viruses

Hepatitis A Virus

HAV belongs to the family *Picornaviridae* and the genus Hepatovirus.[6–8] It has a 7.5 kb positive sense polyadenylated RNA genome with untranslated regions (UTRs) at either termini.[9] The 5′-UTR contains an internal ribosome entry site (IRES) for cap independent translation and is covalently attached to the viral protein 'g' (VPg) (Fig. 1A). A single open reading frame encodes a polyprotein of 252 kDa that is processed to yield structural and nonstructural proteins. The P1 region located at the amino-terminal end encodes the structural proteins VP1, VP2, VP3 and VP4, while the P2 and P3 regions encode the nonstructural proteins associated with viral genome replication and modulation of the host cell.

Hepatitis B Virus

HBV is a double-stranded DNA virus and a member of the family *Hepadnaviridae*. The viral genome is approximately 3.2 kb in size and has an unusual structure comprising two linear strands of DNA

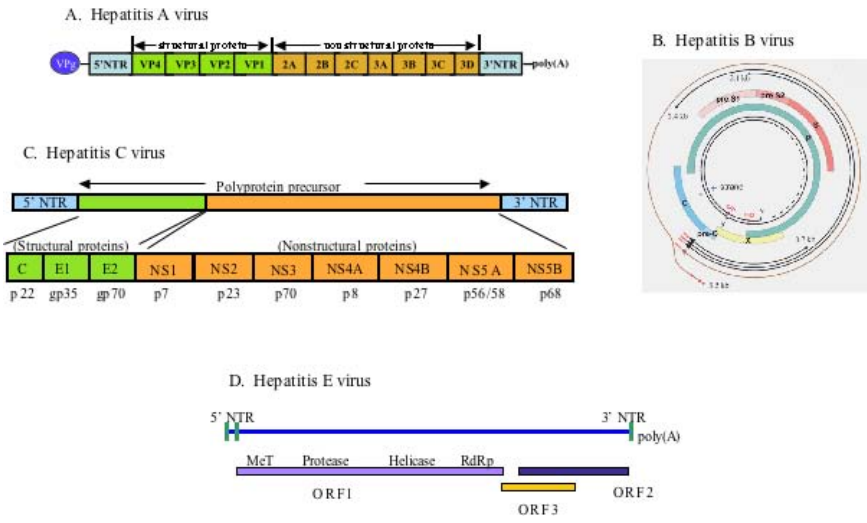


Fig. 1. The genomic organization of hepatitis viruses.

held in a circular configuration by base pairing at the 5′-ends. It codes for four overlapping reading frames (ORFs) required for expression of seven different proteins. The S and C ORFs encode the surface (envelope) and nucleocapsid (core) proteins, respectively (Fig. 1B). The 'P' ORF codes for the viral polymerase and the 'X' ORF codes for the regulatory HBx protein. The surface protein is of three types viz.: small (HBsAg), middle (MHBsAg) and large (LHBsAg). While HBsAg is the major constituent of all HBV particles and is manufactured in large quantities, the other forms are also present in the virus particles in varying amounts. The LHBsAg appears to bind the receptor present on liver cells.

Hepatitis C Virus

HCV belongs to the family *Flaviviridae* and has a positive sense RNA genome of approximately 9500 nucleotides. The whole genome encodes a large polyprotein of about 3000 amino acids. When cleaved by viral and host proteases the HCV polyprotein produces 10 different polypeptides with various biochemical and structural functions.

The 5′ UTR of 324–341 nucleotides contains an IRES element that is required for the cap-independent translation of the viral genomic RNA; a 3′-UTR of about 200 nucleotides is required for viral genome replication. The coding region encodes the core, envelope 1 and 2 structural proteins, and the p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B nonstructural proteins (Fig. 1c). The core protein can bind to RNA to encapsidate the HCV genome. The envelope proteins E1 and E2 form a heterodimer that makes up the viral envelope. The NS3 protein has serine protease and helicase activities, while the NS4A protein functions as a cofactor for the NS3 protease activity.[10] The function of NS4B is unknown. The NS5 region encodes NS5A and NS5B, of which the former is believed to play a role in RNA replication, while the latter has an RNA-dependent RNA polymerase activity.[11] The NS5A protein also has an interferon sensitivity determination region (ISDR), and mutations within this region determine the efficacy of interferon α therapy for HCV infection.

Hepatitis D Virus

Hepatitis D virus (HDV) has a 1700-nucleotide single-stranded RNA genome that encodes the only HDV protein, called the hepatitis delta antigen (HDAg). This protein is expressed in large and small forms, and these are involved in structural as well as regulatory functions. HDV is a replication defective virus that requires the HBV helper function in the form of HBsAg to envelope the HDAg/HDV RNA core particle. The HDV envelope contains all three forms (S, M, and L) of HBsAg, with a predominance of the major (small) form of HBsAg. The nucleocapsid structure of HDV is composed of the viral single-stranded RNA genome and about 60 copies of delta antigen in its large and small forms.

Hepatitis E Virus

Hepatitis E virus (HEV) is a spherical, non-enveloped, RNA virus with a positive sense RNA genome of ~7.2 kb that encodes three overlapping open reading frames (ORFs). ORF1 (~5 kb) encodes the

viral nonstructural polyprotein, ORF2 (~2 kb) encodes the viral major capsid protein, and ORF3 encodes a small protein of undefined function (Fig. 1d). The nonstructural polyprotein ORF1 includes a putative viral methyltransferase (MeT), a papain-like cysteine protease, an RNA helicase and a viral RNA-dependent RNA polymerase (RdRp). The genome contains short 5′-and 3′-NTRs that have the potential to fold into conserved stem–loop and hairpin structures. Such secondary structures are postulated to be important for HEV RNA replication.

The host generates two types of immune responses to control virus infection. The first is a rapid-onset "innate" response against the virus, which primarily involves the synthesis of interferons (IFN) and the stimulation of natural killer lymphocytes. In many cases, the innate response is sufficient to clear the infection. However, if the infection proceeds beyond the first few rounds of viral replication, the adaptive immune response develops antigen-specific antibodies to neutralize extracellular virions and antigen-specific cytotoxic T cells to clear virus-infected cells. Another critical host response following viral infection is the programmed death (apoptosis) of infected cells. To overcome the host immune responses, viruses use various strategies, many of which depend on the modulation of signaling pathways that themselves form the basis of these host responses. Viruses also actively promote the survival of infected cells by regulating critical signaling pathways. In the following sections we will discuss known mechanisms and pathways involved in interferon response, apoptosis and cell survival, and how hepatitis viruses modulate these to establish and propagate infection.

The Interferon System

The interferon (IFN) system is a vital component of the innate immune response. Interferons are small secretory proteins of the cytokine family that help to establish an antiviral state.[12] This system includes two kinds of cells: those that synthesize interferon in response to viral infection, and those that include the neighboring bystander cells that respond to interferon and aid in establishing an antiviral state (Fig. 2). Type I interferons, IFNα and IFNβ, are produced as a direct response to viral infection by leukocytes and fibroblasts, respectively. Together,

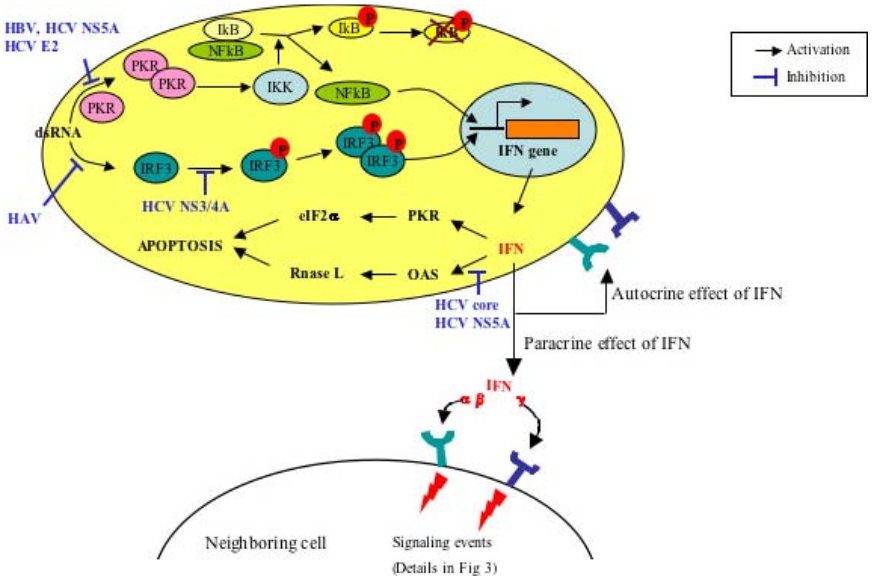


Fig. 2. Interferon action and the effects of hepatitis viruses.

IFNα and β induce the transcriptional activation of at least 30 cellular genes of the cytokine family.[13] Type II interferon, IFNγ, is produced by natural killer (NK) cells and activated T lymphocytes in response to viral infection. Several interferon-induced cellular proteins have been implicated in its antiviral action within an infected cell. These are a dsRNA-induced protein kinase (PKR),[14] the 2′–5′-oligoadenylate synthetase (OAS),[15] the RNA specific adenosine deaminase (ADAR),[16,17] and the Mx protein GTPases.[18] We will consider the known effects of hepatitis viruses on each of these pathways separately. For the sake of clarity, we will divide the interferon system into events that take place inside infected cells and those that result from the action of interferons on uninfected neighboring cells (Fig. 2).

Events in Virus-infected Cells

Most viruses produce double-stranded RNA (dsRNA) at some stage of their replication cycle[12] and these are potent inducers of the IFN

pathway. The dsRNA-dependent serine/threonine protein kinase PKR is a key IFN-inducible protein. Two motifs in amino terminus of PKR bind dsRNA leading to its activation, dimerization and autophosphorylation. Activated PKR catalyzes the phosphorylation of several substrate proteins, the best characterized being the α1 subunit of the eukaryotic protein synthesis initiation factor 2 (eIF2α). Phosphorylation of eIF2α increases its affinity for GDP, thus inhibiting the GDP-GTP exchange that is necessary for continuing translation. This shuts down host as well as viral protein synthesis,[19] leading to death of the infected cell.

Hepatitis viruses have developed several strategies to counter the effects of PKR. The HCV NS5A protein prevents PKR dimerization and activation,[20] thereby preventing PKR-mediated translation inhibition[20–22] (Fig. 2). Exogenously expressed NS5A was found to confer interferon resistance to vesicular stomatitis virus (VSV), which is otherwise sensitive to interferon.[23,24] Mutations within the NS5A region termed as the IFN sensitivity-determining region (ISDR) disrupt the ability of NS5A to bind PKR and inhibit the catalytic activity of PKR. The HCV envelope E2 protein contains a sequence identical to the phosphorylation sites of PKR and eIF2α. As a result, the E2 protein can bind and disrupt PKR dimerization, and inhibit its catalytic activity.[25] These combined effects of NS5A and E2 on PKR activation are proposed to be responsible for HCV persistence.[26–29] A reduction in PKR levels has also been observed in HCC tumor tissues of HBV patients, suggesting a possible role of PKR in promoting tumor growth.[30]

The activation of PKR also affects the status of Nuclear Factor kappa B (NFκB), a member of the Rel family of proteins and a transcription factor central to the expression of various immune regulatory genes.[31] In a quiescent cell, NFκB is present in the cytoplasm in an inactive state as a p65/p50 heterodimer bound to the inhibitor of NFκB (IκB) protein. On phosphorylation of IκB, the complex dissociates, allowing the transport of NFκB into the nucleus, where it regulates the expression of multiple genes. Activated PKR leads to the phosphorylation of IκB,[32] possibly through phosphorylation of its upstream IκB kinase (IKK).[33] In the nucleus, NFκB binds to the IFNγ

promoter and is part of the enhanceasome that contains multiple transcription factors and the IFN regulatory factor (IRF) proteins (Fig. 2).

The IRF proteins are a family of transcription factors (IRF 1–9) that regulate the expression of several genes transcriptionally responsive to IFN.[34–36] Within this family, IRF3 and IRF7 have been identified as key regulators of IFNα/β induction.[37] IRF3 is constitutively expressed in many cell types and is localized to the cytoplasm of normal uninfected cells. Upon virus infection or in the presence of dsRNA, IRF3 is activated by phosphorylation in its C-terminus by a virus-activated kinase (VAK). Some of the components of VAK include IKK-related kinase, IKKε and a TANK binding kinase (TBK1)[38] (Fig. 2). Phosphorylation of IRF3 causes its dimerization and translocation to the nucleus, where it binds regulatory elements leading to the transcription of the genes for IFNα and β, the chemokine RANTES and other proteins that inhibit viral infection. Viral infection also stimulates IRF7 synthesis. Phosphorylation of IRF7 by VAK causes its translocation to the nucleus, where it forms a heterodimer with IRF3 and activates the expression of IFNα genes. Thus, together NFκB and IRFs act as potent stimulators of interferons and several other cytokines that can mount an inflammatory response and contribute to the clearance of viral infection. Understandably then, several viruses have targeted these pathways to establish infection in the host.

The HCV encoded NS3/4A serine protease inhibits the phosphorylation of IRF3,[39] thereby preventing its nuclear translocation and downstream functions (Fig. 2). Expression of the HCV polypeptide or the presence of replicating HCV genome prevents induction of the IRF3 pathway by Sendai virus that is otherwise a potent activator of IRF3.[39] The serine protease activity of NS3/4A is required for this inhibition, suggesting that the protease may target upstream activators like VAK for proteolytic degradation. Further, reporter assays have shown the baseline activity of the interferon stimulated response element (ISRE) to be significantly lower in cells harboring the HCV replicon compared to control cells.[40] The expression of IRF1 was also found to be significantly lower in cells expressing the replicon.[40] The NS5A protein blocks the phosphorylation/activation of IRF1 and the induction of IRF1-dependent promoter activation.[41–43] Noncytopathogenic strains

of HAV inhibit dsRNA-induced IFNβ gene expression as well as apoptosis,[44] while cytopathogenic HAV strains induce apoptosis.[45] The dominant property of noncytopathogenic strains preserves the sites of viral replication for a longer time, thus allowing the establishment of viral infection. The HBV enhancer I/X gene promoter was found to contain a functional interferon-stimulated regulatory element (ISRE) that was responsive to IRF1 and IRF7; however, IFNα and the IRFs showed minimal effects on HBV transcription and replication in the context of the whole genome in cell culture.[46]

Another major pathway of interferon action is the induction of 2′,5′-oligoA synthetase (OAS) resulting in synthesis of 2′ and 5′ linked oligoadenylate molecules. This leads to the activation of RNase L and degradation of cellular RNA in infected cells. Surprisingly, only minimal effects of hepatitis viruses have been observed on this pathway of interferon action. While the HCV core protein was shown to activate OAS transcription[47] in HCV patients undergoing IFNα therapy, the NS5A protein appeared to affect OAS activity.[48] Microarray analysis of IFN-induced genes in liver cells expressing the HCV NS5A protein showed reduced induction of the OAS p69 gene, an effect that appeared to be mediated by the cytokine IL8.[49]

The Mx family includes interferon-inducible proteins with antiviral activity. These proteins are GTPases that belong to the dynamin-like GTPase superfamily.[50,51] IFN-α and β, but not IFN-γ, induce the Mx family of proteins. The spectrum of antiviral activities of the Mx proteins and the molecular mechanisms by which they inhibit viral replication depend on the specific Mx protein and its subcellular localization. The antiviral effects of MxA are dependent upon its oligomerization, GTP binding and hydrolysis.[52] Chronic HBV infection is associated with defective virus particles generated as a result of alternatively spliced HBV RNA.[53] This leads to cytoplasmic accumulation of the viral capsid protein. The MxA protein normally accumulates in the cytoplasm of IFN-treated cells and blocks RNA synthesis. Expression of defective capsid protein has a suppressive effect on the MxA promoter, thereby reducing the antiviral effect of IFN. The HCV core protein down-regulates the MxA gene promoter, leading to decreased MxA levels.[54]

Events in Uninfected Cells

Interferons also provide a warning system to uninfected cells in the vicinity of infected ones and prepare the former to counter an impending viral infection. These events in uninfected cells are based on an interferon-signaling pathway (Fig. 3). The binding of IFN to its cognate receptor results in activation of the JAK/STAT signaling pathway, culminating in the activation of several IFN-α/β-inducible genes. In an uninduced state, the cytoplasmic domains of type 1 IFN receptors, IFNAR-1 and IFNAR-2, remain associated with Janus tyrosine kinases Jak1 and Tyk2.[55,56] The binding to IFN-α/β to these receptors causes the phosphorylation and activation of Jak1 and Tyk2; these activated kinases in turn phosphorylate tyrosine residues in the cytoplasmic domains of the receptor. These phosphotyrosines act as binding sites for the STAT (signal transducers and activators of transcription) proteins through their src homology 2 (SH2) domains.

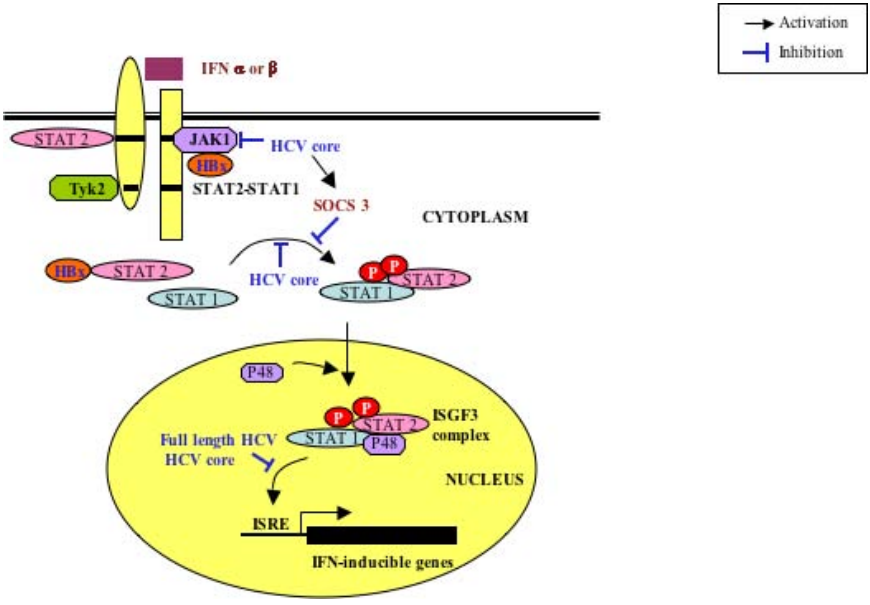


Fig. 3. The JAK-STAT interferon signaling pathway and the effects of hepatitis viruses.

Within this complex, the STAT proteins (STAT1 and STAT2) undergo tyrosine phosphorylation, and this acts as a signal for their dimerization through complementary pTyr/SH2 interactions. These STAT1 and STAT2 homo- and heterodimers translocate to the nucleus and, together with a protein called p48 (IRF-9), form a multimeric transcription factor called ISGF3. In turn, ISGF3 binds to the IFN stimulated response element (ISRE) within the promoters of IFN-stimulated genes (ISGs) and activates their transcription. Stimulation with IL-6 also leads to the phosphorylation of JAK1, JAK2 and STAT3, followed by nuclear translocation of the STAT3 homodimer. As expected, cytokine signaling through the Jak/STAT pathway is tightly controlled by various mechanisms that involve inactivation or removal of activated STAT molecules.[57] Induction of SOCS (suppressors of cytokine signaling) also negatively regulates this pathway.[58–60] The SOCS proteins bind to Jak family tyrosine kinases, resulting in the inhibition of tyrosine phosphorylation of Janus kinases and STAT factors.[57]

A number of viruses are known to modulate the JAK-STAT signaling pathway by using various mechanisms (Goodbourn 2000). Expression of the full-length HCV ORF suppresses the IFN-mediated STAT/DNA binding activity.[61] The HCV core protein binds JAK and has inhibitory effects on the JAK/STAT pathway. Constitutive expression of the HCV core protein reduces tyrosine phosphorylation of the JAK and STAT proteins, thereby modifying STAT-mediated transcription of ISGs. Point or deletion mutants of the HCV core protein, which do not bind the JAK proteins, show no inhibitory effects on the JAK-STAT pathway following IL-6 stimulation of murine CL2 and human HepG2 cells.[62] The HCV core protein further induces the expression of SOCS3, resulting in the inhibition of STAT1-tyrosine phosphorylation.[63] Thus, HCV appears to have developed multiple ways to interfere with the IFN-mediated host defense systems.[62,64]

While the p48 subunit of ISGF3 is reported to be responsible for the IFNα-mediated suppression of HBV expression,[65] the X protein of HBV (HBx) interacts with Jak1 and activates the JAK-STAT signaling pathway.[66] There is sufficient evidence to indicate that the Jak/STAT pathway is involved in cell proliferation and oncogenesis[67]; this

may contribute to the transforming potential of HBx. A crosstalk between the Jak/STAT pathway and the oncogenic Ras pathway has also been reported.[68] Further, activated Jak phosphorylates Shc or Raf, which in turn activates the Ras-Raf-mitogen-activated protein kinase pathway. Thus, HBx-mediated signaling through the Jak/STAT pathway may be one of the factors contributing to the development of hepatocellular carcinoma.

Regulation of Apoptosis

Programmed cell death (PCD), or apoptosis, is a highly controlled physiological process by which damaged cells or cells that are toxic to the host are removed to preserve the integrity of the host organism. It is also critical for several biological processes like embryonic development, host defense mechanisms and immunity. Initiation of apoptosis is an innate response that rapidly and efficiently removes virus-infected cells, thus limiting viral replication and spread. There are at least three pathways leading to apoptotic cell death. One depends on external death signals such as TNFα and Fas ligand, being mediated by the TNF receptor and Fas, respectively.[69] The second pathway is stimulated by internal signals such as oxidative stress, loss of mitochondrial electrostatic potential, virus replication, etc. These internal signals trigger a cascade of events that involve both pro-apoptotic and anti-apoptotic members of the Bcl2 family of proteins. The release of cytochrome c from mitochondria is a crucial event in this process. Both of these pathways activate caspases, a family of proteases that provide the effector function in cell death by degrading cellular proteins. The third pathway is independent of caspases and instead uses the apoptosis inducing factor (AIF). In a dying cell, AIF is released from the mitochondria and is targeted to the nucleus, where it binds DNA and stimulates its degradation.[70]

Signaling Through the TNF Receptor Family

Hepatocyte apoptosis mediated by TNFα is implicated in a wide range of liver diseases including viral hepatitis, non-alcoholic fatty

liver disease and fulminant hepatic failure. TNFα is a pro-inflammatory cytokine produced mainly by activated macrophages and in smaller amounts by other cell types. Hepatocyte apoptosis induced through the extrinsic pathway is initiated by the activation and signaling through the TNF receptor family of proteins following the binding of ligands such as TNFα, Fas ligand (FasL), TNFα-related apoptosis-inducing ligand (TRAIL; also called APO-2L) or APO-3L to their respective receptors. TNFα and FasL are considered to play major roles in liver injury. Following the binding of TNFα to its receptor TNFR1, the receptor trimerizes and together with the adaptor protein TNFR-associated death domain (TRADD) recruits the receptor interacting protein (RIP) and TNFR associated factor (TRAF2) that trigger downstream signaling and activation of NFκB.[71] TRADD can also recruit the Fas-associated death domain (FADD) protein, triggering a series of events leading to FADD-mediated apoptosis (Fig. 4). Binding of FasL to Fas (or CD95) initiates signaling through DISC (death inducing signaling complex) by recruiting FADD to the cytoplasmic tail of Fas. The N-terminal death effector domain (DED) of FADD recruits pro-caspase 8/10, causing its proteolytic activation by the FADD-like IL-1β converting enzyme (FLICE).[72] The death receptor-initiated pathway diverges at this point. One pathway is independent of mitochondria and Bcl2 and is based on the activation of pro-caspase 3[73] (Fig. 4). The other pathway is mediated through caspase 8/10-mediated cleavage of Bid, and the truncated Bid (tBid) molecule opens mitochondrial channels causing the release of cytochrome c.[73] An alternate pathway in which TRAF2 or the Fas-binding protein Daxx recruits the Ask1 kinase results in activation of the c-Jun N-terminal kinase (JNK) and suppression of the anti-apoptotic effects of Bcl2 on mitochondria. Depending upon their ability to activate caspase 8, caspase10 or NFκB, the TNF family members are regarded as pro-apoptotic or anti-apoptotic proteins.

The HCV core protein activates NFκB signaling by binding to the death domain of TNFR1,[74,75] thereby contributing to liver damage. However, expression of the HCV core protein confers resistance to TNF-induced apoptosis. Immunohistochemical studies showed a higher frequency of NFκB staining in HCV-infected livers compared

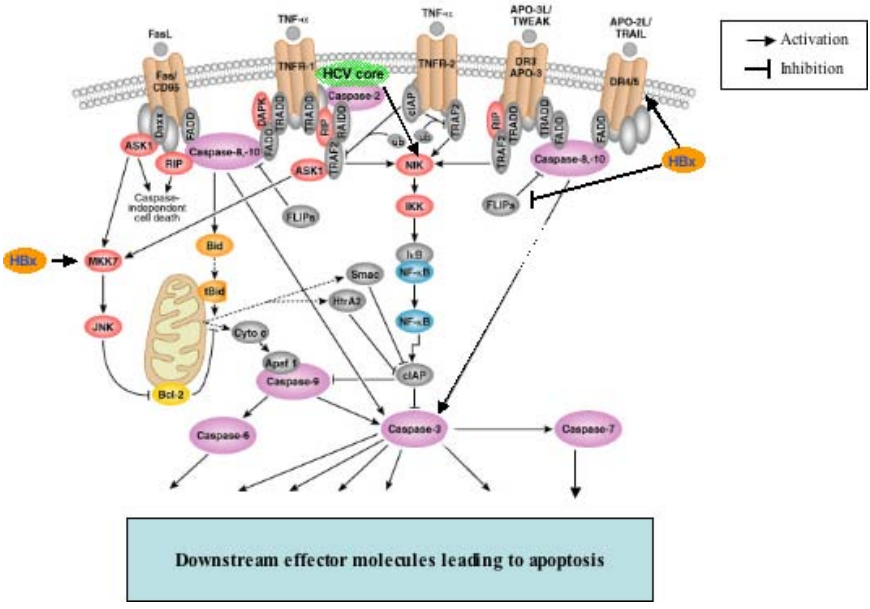


Fig. 4. Apoptosis pathways and their regulation by hepatitis viruses.

to controls.[76] Further, mobility shift assays also showed higher NFκB activity in HCV core-expressing cells.[77] Signaling through TNFR is variable and highly controversial, ranging from being protective, irrelevant or pro-apoptotic. The outcome of the signal is decided by the stimulus and the cell type. While HCV can cause apoptotic death by signaling through TNFR, it may also overcome the TNF-mediated death pathway by activation of NFκB (Fig. 4) and contribute to viral persistence and hepatocarcinogenesis.[76,77] The anti-apoptotic effects of the HCV core protein probably account for a chronically activated and persistent state in HCV infected cells, leading to hepatocarcinogenesis. The role of TNFR and Fas signaling in hepatitis B is generally linked to sensitization of HBV-infected hepatocytes to TNF or Fas ligand. The HBx protein was shown to induce apoptosis by prolonged stimulation of the stress-mediated MEKK1 pathway[78] and by up-regulating expression of the TRAIL receptor DR5.[79] It was also shown to bind to the cellular FLICE inhibitor protein

(c-FLIP), a key regulator of the death-inducing signaling complex, and to suppress its anti-apoptotic effects.[80] In chronic and fulminant hepatitis B patients the Fas/FasL system has also been shown to be important for liver injury.[81,82] By sensitizing liver cells to TNF and FasL, HBV promotes liver injury, an event that precedes cirrhosis and hepatocarcinogenesis. Further, since TNF sensitization can be overcome by mitogenic factors, this may also be a mechanism to select for neoplastic hepatocytes that survive by synthesizing large amounts of growth factors.

Endoplasmic reticulum stress

The endoplasmic reticulum (ER) is sensitive to a variety of cellular stresses, including viral infection, alterations in calcium levels, inhibition of protein glycosylation, etc.[83] Excessive protein synthesis and accumulation of unfolded or misfolded proteins lead to ER stress triggering the unfolded protein response (UPR) and protein degradation. This signaling pathway is mediated by three ER resident sensor proteins: the type-I ER transmembrane protein kinase IRE-1, the activating transcription factor 6 (ATF-6) and the PKR-like ER kinase (PERK). In the absence of ER stress, the ER resident protein GRP78 is associated with each of the sensors.[84–86] Upon accumulation of unfolded/misfolded proteins, GRP78 dissociates from its partners, activating the different pathways (Fig. 5).

Subgenomic replicons of HCV as well as expression of the NS5A protein alone induces ER stress, as observed by the appearance of a transcriptionally active form of the ATF protein, ATF6α(N). This translocates to the nucleus, where it transcriptionally activates ER chaperones.[84–86] HCV compels cells to adapt to ER stress by different mechanisms. Cells harboring HCV replicons have lower levels of eIF2α phosphorylation, indicating that the overall protein synthesis is higher.[87] HCV E2 modulates global translation by inhibition of IFN-induced PKR as well as by binding and inhibiting PERK. HCV infection can suppress the degradation of misfolded proteins and simultaneously stimulate the synthesis of the viral protein in the ER. The IRE1-XBP1 (X-box binding protein) pathway directs both protein refolding

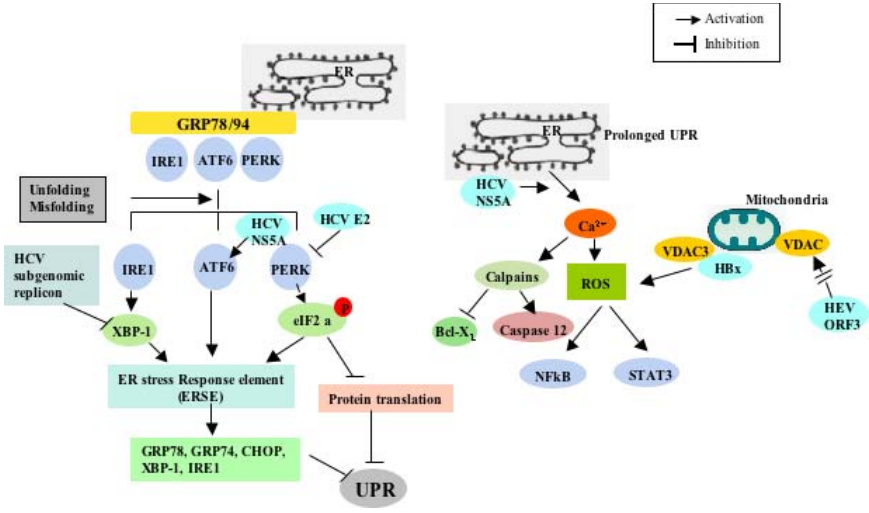


Fig. 5. The effects of hepatitis viruses on ER stress and ROS generation.

and degradation in response to ER stress. In cells carrying the HCV subgenomic replicons, XBP1 expression is elevated but its transactivating activity is repressed.[88] This prevents the induction of EDEM (ER degradation enhancing mannosidase-like protein) that is required for degrading misfolded proteins[87] (Fig. 5). Thus, cells expressing HCV replicons suppress the degradation of misfolded proteins, including viral proteins, and may contribute to viral persistence. The HBV overcomes ER stress by differentially regulating expression of the three forms of its surface envelope protein. The large, middle and the small (major) forms of the surface protein are cotranslationally inserted into the ER, which together with the nucleocapsid particles forms mature virions that are secreted. The small and the middle forms can be secreted in the absence of other viral proteins. However, the large surface protein (LS) cannot be secreted when expressed alone. When LS is expressed in large amounts, its retention affects the intracellular ER stress response, leading to a disruption in the viral life cycle.[89] The virus has developed a feedback mechanism whereby LS can transactivate the S promoter, thus increasing the expression of middle and small forms of HBsAg. This again balances the ratio of the

three forms and restores virus particle formation.[90] The hepatitis E virus capsid (ORF2) protein has recently been shown to use the ER associated degradation (ERAD) pathway to retrotranslocate from the ER back into the cytoplasm to interfere with IκB ubiquitination as a means to regulate NFκB activity.[91] The functional consequences of this on HEV pathogenesis remain to be understood.

Reactive oxygen species (ROS)

Cellular oxidative stress is characterized by an increase in the intracellular levels of reactive oxygen species (ROS) that can affect cell survival as well as death pathways. ROS can function as a key player in different signaling cascades by activating transcription factors such as NFκB and STAT3[92,93] that are involved in a wide variety of cellular functions ranging from cellular growth, proliferation, DNA replication and repair to functions involved in cell death and cancer development.[94] ER stress can release calcium from intracellular ER stores into the cytosol, and this can trigger oxidative stress, activate NFκB and modulate the mitochondrial death pathway (Fig. 5). Increased intracellular calcium can activate calpains that cleave and inactivate the anti-apoptotic Bcl-X_L protein, and can also activate caspase 12. Increased production of ROS has been observed in the livers of chronically infected HBV and HCV patients. The HBx protein induces oxidative stress and elevates ROS production by its direct association with the voltage-dependent anion channel 3 (VDAC3) protein in the mitochondrial outer membrane. This interaction leads to a decrease in the mitochondrial membrane potential and elevates ROS.[95] A mutant HBx not able to translocate to mitochondria but capable of binding to VDAC3 failed to elevate ROS levels.[94] In turn, HBx levels were also found to be regulated by cellular ROS levels.[96] Therefore, intracellular microenvironments generating ROS, such as severe inflammation, may aggravate the pathogenesis of liver disease by accumulating the HBx protein. The HCV NS5A protein also induces ER stress, causes alteration of calcium homeostasis, and triggers ROS elevation.[97] These events result in activation and translocation of transcription factors including STAT3 and NFκB to the nucleus. The HEV ORF3

protein has been found to up-regulate expression of the mitochondrial outer membrane protein VDAC1 and to increase the mitochondrial transmembrane potential.[98] This viral protein also provides protection against apoptotic signals, though the mechanism still remains to be fully elucidated.

Cell Survival Strategies

Besides inhibiting the innate apoptotic response of the host, many viruses actively stimulate signaling pathways that promote survival and proliferation of the infected cell. This aids in viral replication and helps promote the establishment and persistence of viral infection.

Regulation of the MAP Kinase Pathway

The MAPK signaling pathway is initiated in response to external or internal stimuli such as growth factors, stress, DNA damage, virus infection, etc. In vertebrates, three MAPK pathways have been characterized.[99] These are the extracellular signal-regulated kinases (MAPK/ERK), the p38 MAPK and the stress-activated protein kinase/c-jun (SAPK/JNK). All these pathways are based on a three-kinase signaling module consisting of a MAPK that is activated by an upstream MAPK kinase (MAPKK), which itself is activated by a MAPK kinase kinase (MAPKKK). The prototype of the MAPK signaling pathway is the Raf/MEK/ERK cascade, a major pathway for cell growth and differentiation. Signaling through this pathway activates ERK; activated ERK1/2 translocates to the nucleus to activate a variety of transcription factors such as Elk, c-Fos, c-Jun, STAT3, etc., that can then regulate cell proliferation and differentiation.

Many viruses activate ERK in a bid to regulate cell survival, proliferation and differentiation, and hepatitis viruses are no exception. HBV activates epidermal growth factor receptor (EGFR) mediated activation of the Raf/MEK/ERK pathway. On binding to EGF, the cytoplasmic domain of EGFR undergoes autophosphorylation, leading to its interaction with adaptor proteins Shc and Grb2 that mediate Raf/MEK/ERK activation[100] (Fig. 6). Further, HBx transactivates the

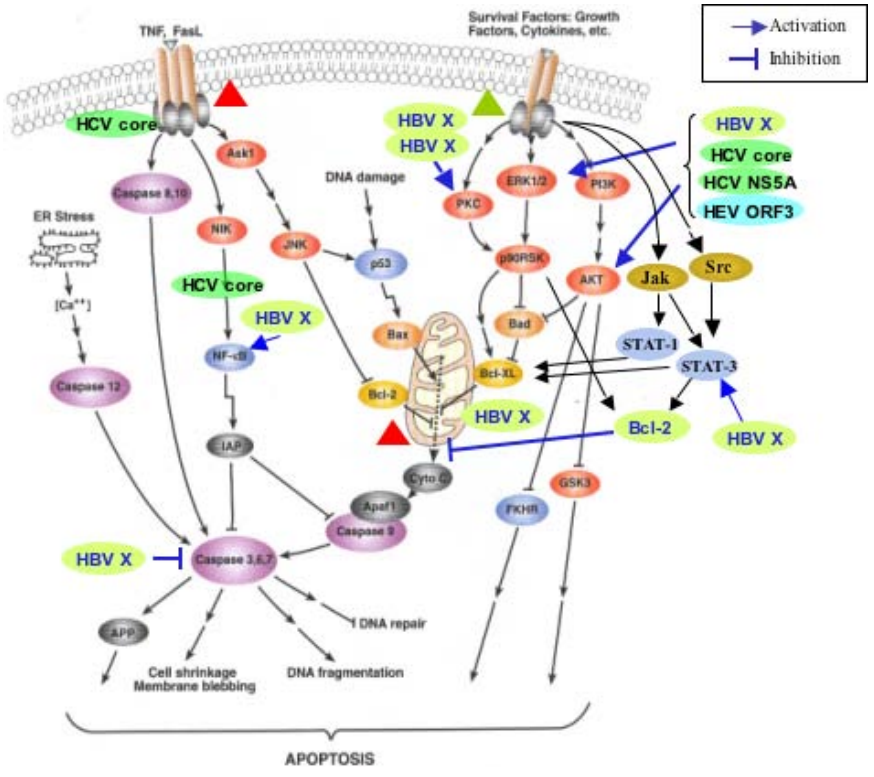


Fig. 6. An overview of the survival and death pathways and their modulation by hepatitis viruses.

EGFR promoter, up-regulating its expression and creating a positive feedback loop that promotes cell growth and survival.[101] The truncated middle surface antigen of HBV (mHBs) also activates this pathway, but in a PKC-dependent manner. The HBx protein has also been shown to stabilize the Ras-GTP complex, again promoting downstream signaling through Raf and MAPKs.[102]

The HCV core protein binds to a Raf inhibitory protein 14-3-3 to relieve its repression of the Raf/MEK/ERK pathway.[103] Cell lines expressing the HCV core protein also show higher activities of Raf, MEK and Elk1, and a higher degree of phosphorylation and activation of ERK.[104] The HCV NS5A protein has conserved PXXP motifs that

bind to the src homology 3 (SH3) domains found in a diverse group of signal-transducing molecules,[105] including Grb2[106] and the p85 subunit of PI3K[107] leading to subversion of signaling through the Raf/MEK/ERK and PI3K/Akt pathways. The NS5A-Grb2 interaction also results in down-regulation of IFN expression,[108] while its interaction with PKR leads to inhibition of PKR, an effector of the interferon pathway. Thus HCV adopts a dual strategy for survival by activating MAPK on one hand and inhibiting the interferon response on the other. The HEV ORF3 protein also has PXXP motifs and binds to SH3 domains of multiple proteins, including Src, Grb2, Hck, Fyn, PLC-γ and the p85 regulatory subunit of PI3K[109]; these proteins are upstream modulators of Raf/MEK/ERK, PI3K/Akt and PLC-γ/PKC signaling pathways. Increased activity of ERK has been observed in ORF3 expressing cells and this was shown to be due to its binding to and inhibition of a MAPK phosphatase.[110]

Regulation of the PI3 Kinase/Akt (PKB) Pathway

On binding their cognate cell surface receptors, extracellular signals like growth factors recruit a large number of cytoplasmic signaling molecules, phosphatidylinositol-3-kinase (PI3K) being one of them.[111,112] Recruitment of PI3K to the cell membrane by growth factors or other docking molecules activates downstream effectors such as the kinase PDK1 and its substrate Akt (also called protein kinase B, PKB). By recruiting and activating signaling complexes, PI3K triggers complex cellular events such as cell cycle entry, cell survival and migration. Akt is a serine/threonine kinase that on activation phosphorylates a number of downstream effectors including other kinases and transcription factors. Activation of Akt results in the phosphorylation and inhibition of pro-apoptotic signals from proteins such as BAD,[113] caspase 9,[114] and Forkhead,[115,116] while promoting cell survival signals mediated by NFκB[117] (Fig. 6). In its non-phosphorylated state, BAD translocates from the cytosol to the mitochondria and promotes apoptotic death by binding and inhibiting anti-apoptotic proteins Bcl2 and Bcl-X_L. Activated Akt phosphorylates BAD, retaining it in the cytosol and thereby preventing

apoptosis. Many viruses encode proteins that target the PI3K/Akt pathway because of the central role it plays in regulating host cell survival. The HCV NS5A protein associates with the p85 regulatory subunit of PI3K to promote the PI3K/Akt-mediated BAD inactivation.[107] The HBx protein also promotes cell survival through this pathway by inactivating caspase 3, which is a central effector molecule in apoptosis. The HEV ORF3 protein also activates Akt and phosphorylation of its downstream substrates (Jameel; unpublished). While this is generally understood to promote cell survival, the mechanistic details are not yet clear.

Regulation of the NFκB Pathway

Transcription factor NFκB induces the expression of multiple genes whose products function in a variety of pathways ranging from cell survival and proliferation to the regulation of immune and inflammatory responses.[31] NFκB is activated in response to various stimuli, and the signal transduction pathways involve different signaling proteins such as the NFκB-inducing kinase (NIK), mitogen-activated protein kinase/extracellular signal-regulated kinase kinase 1 (MEKK1), IL-1 receptor-associated kinase (IRAKs), protein kinase C (PKC), transforming growth factor β-activated kinase (TAK1), TNF receptor-associated factor (TRAF) and many others that converge centrally on the IκB kinase (IKK) signaling complex.[118] NFκB plays a crucial role in control of cell proliferation and survival by inducing the expression of cyclin D1, c-myc and other positive regulators of the cell cycle. It also activates the expression of anti-apoptotic genes whose products include members of the Bcl2 family, such as Bcl-X_L, A1, Bfl1, and cellular inhibitors of apoptosis (c-IAP1, c-IAP2 and XIAP), TRAF1, TRAF2 and FLICE-inhibitory protein cFLICE. Different viruses use this dynamic pathway to prolong the survival of host cells to enable the virus to complete its replication cycle and establish infection in the host. The HBx protein is a potent inducer of NFκB and is believed to play a critical role in hepatocarcinogenesis during chronic HBV infection.[119] HBx activates NFκB by MEKK-1 activation mediated by the Ras and the PKC pathways. The expression of HCV proteins is

generally correlated with NFκB activation, and this has been suggested as a mechanism for evading apoptosis of infected cells.[77] It is generally agreed that the HCV NS5A protein activates the NFκB pathway.[120] by phosphorylation of IκBα and its degradation by calpain[121] (Fig. 5). The role of the HCV core protein is controversial, with some reports suggesting NFκB activation[122–124] and others reporting NFκB down-regulation.[125,126] It is likely that the outcome is dependent upon the expression system and cell lines used for the different studies. In transfected cells expressing the HEV ORF3 protein, increased nuclear translocation of NFκB p65 has been observed (Jameel; unpublished), but the mechanistic details and downstream consequences are unclear.

Conclusion

Viruses have evolved complex and effective mechanisms to circumvent the host immune response. On entering a naïve cell, the virus needs sufficient time to replicate, establish an infection, proliferate and spread the infection. The fate of a viral infection is determined by a critical balance between the host immune response and various strategies adopted by viruses to evade those responses. Viruses have co-evolved with their hosts and have acquired the capacity to target critical steps in host cell signaling. Most often this involves hijacking or mimicking one or more host cell proteins that control critical cellular pathways to promote virus replication and the infection process. For example, because of the central role of PKR and NFκB in the cell survival pathway, these have become the primary targets of numerous viruses. While *in vitro* studies and analyses of signaling pathways have provided us much insight into the molecular nature of host-virus interactions and their consequences, caution must be exercised in over-interpreting the physiological relevance of these results. For many viral systems, these results are based purely on *in vitro* analyses in the absence of other viral proteins or critical host components. This problem is particularly relevant for hepatitis viruses that generally have neither good small animal models nor robust *in vitro* expression systems. The work with hepatitis viruses and host signaling is currently

poised at a point where interesting mechanistic leads have been obtained. These await confirmation in robust *in vivo* systems of virus infection that are slow to develop.

Acknowledgments

We are grateful to Cell Signaling Technology, Beverly, MA, USA for permission to utilize signaling pathway templates for Figs. 4 and 6. A Senior Research Fellowship of The Wellcome Trust, UK to SJ and institutional funds from ICGEB supported research in the authors' laboratory. The award of a Research Fellowship of the CSIR, India to SMM is gratefully acknowledged.

References

1. Zuckerman AJ. (1983) The history of viral hepatitis from antiquity to the present. In F. Deinhardt & J. Deinhardt (eds), *Viral Hepatitis: Laboratory and Clinical Science*, pp. 3–32. Marcel Dekker, New York.
2. Choo QL, Weiner AJ, Overby LR, *et al.* (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**: 359–362.
3. Reyes GR, Purdy MA, Kim JP, *et al.* (1990) Isolation of a cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis. *Science* **247**: 1335–1339.
4. Jameel S. (1999) Molecular biology and pathogenesis of hepatitis E virus. *Exp Rev Mol Med*, http://www.expertreviews.org/99001271h.htm.
5. Chang MH. (1995) Chronic viral hepatitis. *Indian J Pediatr* **62**: 673–679.
6. Matthews RE. (1979) The classification and nomenclature of viruses. *Intervirology* **11**: 133–135.
7. Matthews RE. (1982) Classification and nomenclature of viruses. *Intervirology* **17**: 1–199.
8. Melnick JL. (1992) Properties and classification of hepatitis A virus. *Vaccine* **10 (Suppl. 1)**: S24–S26.
9. Hollinger FB, Emerson SU. (2001) Hepatitis A virus. In D.M. Knipe & P.M. Howley (eds), *Fields Virology, 4th Edition*, pp. 799–840. Lippincott Williams & Wilkins, New York.

10. Failla C, Tomei L, De Francesco R. (1994) Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. *J Virol* **68**: 3753–3756.
11. Behrens SE, Tomei L, De Francesco R. (1996) Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO J* **15**: 12–22.
12. Samuel CE. (2001) Antiviral Actions of Interferons. *Clin Microbiol Rev* **14**: 778–807.
13. Sen GC, Ransohoff RM. (1993) Interferon induced antiviral actions and their regulation. *Adv Virus Res* **42**: 57–102.
14. Clemens MJ, Elia A. (1997) The double stranded RNA-dependent protein kinase PKR: structure and function. *J Interferon Cytokine Res* **17**: 503–524.
15. Player MR, Torrence PF. (1998) The 2-5A system: modulation of viral and cellular processes through acceleration of RNA degradation. *Pharmacol Ther* **78**: 55–113.
16. Patterson JB, Samuel CE. (1995) Expression and regulation by interferon of a double-stranded RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Mol Cell Biol* **15**: 5376–5388.
17. Patterson JB, Thomis DC, Hans SL, Samuel CE. (1995) Mechanism of interferon action: double-stranded RNA-specific adenosine deaminase from human cells is inducible by alpha and gamma interferons. *Virology* **210**: 508–511.
18. Haller O, Frese M, Kochs G. (1998) Mx proteins: mediators of innate resistance to RNA viruses. *Rev Sci Technol Off Int Epizootol* **17**: 220– 230.
19. Kaufman RJ. (1999) Double stranded RNA activated protein kinase mediates virus induced apoptosis: a new role for an old factor. *Proc Natl Acad Sci* **96**: 11693–11695.
20. Gale M Jr, Blakely CM, Kwieciszewski B, *et al.* (1998) Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. *Mol Cell Biol* **18**: 5208–5201.
21. Gale M Jr, Blakely CM, Hopkins DA, *et al.* (1998) Regulation of Interferon-Induced Protein Kinase PKR: Modulation of P58IPK Inhibitory Function by a Novel Protein, P52rIPK. *Mol Cell Biol* **18**: 859–871.
22. Gale M Jr, Korth MJ, Tang NM, *et al.* (1997) Evidence that hepatitis C virus resistance to interferon is mediated through repression of the

PKR protein kinase by the nonstructural 5A protein. *Virology* **14**: 230, 217–227.
23. Gale M Jr, Kwieciszewski B, Dossett M, *et al.* (1999) Antiapoptotic and oncogenic potentials of hepatitis C virus are linked to interferon resistance by viral repression of the PKR protein kinase. *J Virol* **73**: 6506–6516.
24. Gale M Jr, Katze MG. (1998) Molecular mechanisms of interferon resistance mediated by viral-directed inhibition of PKR, the interferon induced protein kinase. *Pharmacol Ther* **78**: 29–46.
25. Taylor DR, Shi ST, Romano PR, *et al.* (1999) Inhibition of the Interferon-inducible Protein Kinase PKR by HCV E2 Protein. *Science* **285**: 107–110.
26. Koromilas AE, Roy S, Barber GN, *et al.* (1992) Malignant transformation by a mutant of the IFN-inducible dsRNA-dependent protein kinase. *Science* **257**: 1685–1689.
27. Barber GN, Thompson S, Lee TG, *et al.* (1994) The 58-kilodalton inhibitor of the interferon-induced double-stranded RNA-activated protein kinase is a tetratricopeptide repeat protein with oncogenic properties. *Proc Natl Acad Sci USA* **91**: 4278–4282.
28. Barber GN, Wambach M, Thompson S, *et al.* (1995) Mutants of the RNA-dependent protein kinase (PKR) lacking double-stranded RNA binding domain I can act as transdominant inhibitors and induce malignant transformation. *Mol Cell Biol* **15**: 3138–3146.
29. Benkirane M, Neuveut C, Chun RF, *et al.* (1997) Oncogenic potential of TAR RNA binding protein TRBP and its regulatory interaction with RNA-dependent protein kinase PKR. *EMBO J* **16**: 611–624.
30. Chen GG, Lai PB, Ho RL, *et al.* (2004) RNA-activated protein kinase in hepatocellular carcinoma associated with hepatitis B virus. *J Med Virol* **73**: 187–194.
31. Ghosh S, May MJ, Kopp EB. (1998) NFκB and Rel proteins; evolutionary conserved mediators of immune response. *Annu Rev Immunol* **16:** 225–260.
32. Maran A, Maitra RK, Kumar A, *et al.* (1994) Blokage of NF-kappa B signaling by selective ablation of an mRNA targeted by 2-5 A antisense chimeras. *Science* **265**: 789–792.
33. Israel A. (2000) The IKK complex: an integrator of all signals that activate NFκB. *Trends Cell Biol* **10:** 129–133.
34. Fujita T, Kimura Y, Miyamoto M, *et al.* (1989) Induction of endogenous IFNa and IFNb genes by a regulatory transcription factor IRF-1. *Nature* **337**: 270–272.

35. Watanabe N, Sakakibara J, Hovanessian AG, *et al.* (1991) Activation of IFNb element by IRF-1 requires a post translational event in addition to IRF-1 synthesis. *Nucleic Acids Res* **19**: 4421–4428.
36. Nguyen H, Hiscott J, Pitha PM. (1997) The growing family of interferon regulatory factors. *Cytokine Growth Factor Rev* **8**: 293–312.
37. Weaver BK, Kumar KP, Reich NC. (1998) Interferon regulatory factor 3 and CREB binding protein/p300 are subunits of double stranded RNA activated transcription factor DRAF1. *Mol Cell Biol* **18**: 1359–1368.
38. Sharma S, Tenoever BR, Grandvaux N, *et al.* (2003) Triggering the interferon antiviral response through an IKK related pathway. *Science* **300**: 1148–1151.
39. Foy E, Li K, Wang C, *et al.* (2003) Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease. *Science* **300**: 1145–1148.
40. Kanazawa N, Kurosaki M, Sakamoto N, *et al.* (2004) Regulation of hepatitis C virus replication by interferon regulatory factor 1. *J Virol* **78**: 9713–9720.
41. Kirchhoff S, Koromilas AE, Schaper F, *et al.* (1995) IRF-1 induced cell growth inhibition and interferon induction requires the activity of the protein kinase PKR. *Oncogene* **11**: 439–445.
42. Kumar A, Haque J, Lacoste J, *et al.* (1994) Double-stranded RNA-dependent protein kinase activates transcription factor NFκB by phosphorylating IκB. *Proc Natl Acad Sci USA* **91**: 6288–6292.
43. Kumar A, Yang Y-L, Flati V, *et al.* (1997) Deficient cytokine signaling in mouse embryo fibroblasts with a targeted deletion in the PKR gene: role of IRF-1 and NF-κB. *EMBO J* **16**: 406–416.
44. Brack K, Berk I, Magulski T, *et al.* (2002) Hepatitis A virus inhibits cellular antiviral defense mechanisms induced by double-stranded RNA. *J Virol* **76**: 11920–11930.
45. Goswami BB, Kulka M, Ngo D, Cebula TA. (2004) Apoptosis induced by a cytopathic hepatitis A virus is dependent on caspase activation following ribosomal RNA degradation but occurs in the absence of 2′-5′ oligoadenylate synthetase. *Antiviral Res* **63**: 153–166.
46. Alcantara FF, Tang H, McLachlan A. (2002) Functional characterization of the interferon regulatory element in the enhancer 1 region of the hepatitis B virus genome. *Nucleic Acids Res* **30**: 2068–2075.
47. Naganuma A, Nozaki A, Tanaka T, *et al.* (2000) Activation of the interferon-inducible 2′-5′-oligoadenylate synthetase gene by hepatitis C virus core protein. *J Virol* **74**: 8744–8750.

48. Murashima S, Kumashiro R, Ide T, *et al.* (2000) Effect of interferon treatment on serum 2′,5′-oligoadenylate synthetase levels in hepatitis C-infected patients. *J Med Virol* **62**: 185–190.
49. Girard S, Shalhoub P, Lescure P, *et al.* (2002) An altered cellular response to interferon and up-regulation of interleukin-8 induced by the hepatitis C viral protein NS5A uncovered by microarray analysis. *Virology* **295**: 272–283.
50. Staeheli P, Pitossi F, Pavlovic J. (1993) Mx proteins: GTPases with antiviral activity. *Trends Cell Biol* **3**: 268–272.
51. Van der Bliek AM. (1999) Functional diversity in the dynamin family. *Trends Cell Biol* **3**: 96–102.
52. Pitossi F, Blank A, Schroder A, *et al.* (1993) A functional GTP-binding motif is necessary for antiviral activity of Mx proteins. *J Virol* **67**: 6726–6732.
53. Rosmorduc O, Petit MA, Pol S, *et al.* (1995) *In vivo* and *in vitro* expression of the defective HBV particles generated by the 2.2 kb singly spliced RNAs. *Hepatol* **22**: 10–19.
54. Lucas DS, Bartolome J, Carreno V. (2005) Hepatitis C virus core protein down-regulates transcription of interferon-induced antiviral genes. *J Infect Dis* **191**: 93–99.
55. Colamonici OR, Yan H, Domanski P, *et al.* (1994) Direct binding and tyrosine phosphorylation of the a subunit of the type 1 IFN receptor by the p135 tyk2 tyrosine kinase. *Mol Cell Biol* **14:** 8133–8142.
56. Novick D, Cohen B, Rubinstein M. (1994) The human interferon alpha/beta receptor: characterization and molecular cloning. *Cell* **77:** 391–400.
57. Hilton DJ. (1999) Negative regulators of cytokine signal transduction. *Cell Mol Life Sci* **55:** 1547–1558.
58. Starr R, Willson TA, Viney EM, *et al.* (1997) A family of Farci cytokine-inducible inhibitors of signaling. *Nature* **387**: 917–921.
59. Endo TA, Masuhara M, Yokouchi M, *et al.* (1997) A new protein containing an SH2 domain that inhibits JAK kinases. *Nature* **387**: 921–924.
60. Naka T, Narazaki M, Hirata M, *et al.* (1997) Structure and function of a new STAT-induced STAT inhibitor. *Nature* **387**: 924–929.
61. Heim MH, Moradpour D, Blum HE. (1999) Expression of Hepatitis C Virus Proteins Inhibits Signal Transduction through the Jak-STAT Pathway. *J Virol* **73**: 8469–8475.
62. Hosui A, Ohkawa K, Ishida H, *et al.* (2003) Hepatitis C virus core protein differently regulates the JAK-STAT signaling pathway under

interleukin-6 and interferon-gamma stimuli. *J Biol Chem* **278**: 28562–28571.
63. Bode JG, Ludwig S, Ehrhardt C, *et al.* (2003) IFN-alpha antagonistic activity of HCV core protein involves induction of suppressor of cytokine signaling-3. *FASEB J* **17**: 488–490.
64. Melen K, Fagerlund R, Nyqvist M, *et al.* (2004) Expression of hepatitis C virus core protein inhibits interferon-induced nuclear import of STATs. *J Med Virol* **73**: 536–547.
65. Nakao K, Nakata K, Yamashita M, *et al.* (1999) p48 (ISGF-3gamma) is involved in interferon-alpha-induced suppression of hepatitis B virus enhancer-1 activity. *J Biol Chem* **274**: 28075–28078.
66. Lee YH, Yun Y. (1998) HBx protein of hepatitis B virus activates Jak1-STAT signaling. *J Biol Chem* **273**: 25510–25515.
67. Bromberg J. (2002) STAT proteins and oncogenesis. *J Clin Invest* **109**: 1139–1142.
68. Kim H, Lee YH, Won J, Yun Y. (2001) Through induction of juxtaposition and tyrosine kinase activity of Jak1, X-gene product of hepatitis B virus stimulates Ras and the transcriptional activation through AP-1, NF-kappaB, and SRE enhancers. *Biochem Biophys Res Commun* **286**: 886–894.
69. Schulze-Osthoff K, Ferrari D, Los M, *et al.* (1998) Apoptosis signaling by death receptors. *Eur J Bio Chem* **254**: 439–459.
70. Hong SJ, Dawson TM, Dawson VL. (2004) Nuclear and mitochondrial conversations in cell death: PARP-1 and AIF signaling. *Trends Pharmacol Sci* **25**: 259–264.
71. Santoro GM, Rossi A, Amici C. (2003) NFkB and virus infection: who controls whom. *EMBO J* **22**: 2552–2560.
72. Kavurma MM, Khachigian LM. (2003) Signaling and transcriptional control of Fas ligand gene expression. *Nature* **10:** 36–44.
73. Budiharjo I, Oliver H, Lutter M, *et al.* (1999) Biochemical pathways of caspase activation during apoptosis. *Annu Rev Cell Dev Biol* 269–290.
74. Zhu N, Khoshan A, Schneider R, *et al.* (1998) Hepatitis C virus core protein binds to the cytoplasmic domain of tumor necrosis factor (TNF) receptor 1 and enhances TNF-induced apoptosis. *J Virol* **72**: 3691–3697.
75. Chen Cm, You LR, Hwang LH, Lee YW. (1997) Direct interaction of Hepatitis C virus core protein with the cellular lymphotoxin β receptor

modulates the signal pathway of the lymphotoxin β receptor. *J Virol* **71:** 9417–9426.
76. Tai DI, Tsai SL, Chang YH, *et al.* (2000) Constitutive activation of nuclear factor kB in hepatocellular carcinoma. *Cancer* **89**: 2274–2281.
77. Tai DI, Tsai SL, Chen YM, *et al.* (2000) Activation of nuclear factor kappaB in hepatitis C virus infection: implications for pathogenesis and hepatocarcinogenesis. *Hepatology* **31**: 656–664.
78. Su F, Schneider RJ. (1997) Hepatitis B virus HBx protein sensitizes cells to apoptotic killing by tumor necrosis factor alpha. *Proc Natl Acad Sci* **94**: 8744–8749.
79. Janssen HL, Higuchi H, Abdulkarim A, Gores GJ. (2003) Hepatitis B virus enhances tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) cytotoxicity by increasing TRAIL-R1/death receptor 4 expression. *J Hepatol* **39**: 414–420.
80. Kim KH, Seong BL. (2003) Pro-apoptotic function of HBV X protein is mediated by interaction with c-FLIP and enhancement of death-inducing signal. *EMBO J* **22**: 2104–2116.
81. Lee JY, Chae DW, Kim SM, *et al.* (2004) Expression of FasL and perforin/granzyme B mRNA in chronic hepatitis B virus infection. *J Viral Hepatol* **11**: 130–135.
82. Rivero M, Crespo J, Fabrega E, *et al.* (2002) Apoptosis mediated by the Fas system in the fulminant hepatitis by hepatitis B virus. *J Viral Hepatol* **9**: 107–113.
83. Ji C, Kaplowitz. (2004) Hyperhomocysteinemia, endoplasmic reticulum stress and alcoholic liver injury. *World J Gasteroenterol* **10**: 1699–1708.
84. Liu CY, Schroder M, Kaufman RJ. (2000) Ligand-independent dimerization activates the stress response kinases IRE1 and PERK in the lumen of the endoplasmic reticulum. *J Biol Chem* **275**: 24881–24885.
85. Liu CY, Wong HN, Schauerte JA, Kaufman RJ. (2002) The protein kinase/endoribonuclease IRE1a that signals the unfolded protein response has a lumenal amino-terminal ligand-independent dimerization domain. *J Biol Chem* **277**: 18346–18356.
86. Shen J, Chen X, Hendershot L, Prywes R. (2002) ER stress regulation of ATF6 localization by dissociation of BiP/GRP78 binding and unmasking of golgi localization signals. *Dev Cell* **3**: 99–111.
87. Tardif KD, Mori K, Siddiqui A. (2002) Hepatitis C virus subgenomic replicons induce endoplasmic reticulum stress activating an intracellular signaling pathway. *J Virol* **76**: 7453–7459.

88. Tardif KD, Mori K, Kaufman RJ, Siddiqui A. (2004) Hepatitis C virus suppresses the IRE1-XBP1 pathway of the unfolded protein response. *J Biol Chem* **279**: 17158–17164.
89. Khan N, Guarnieri M, Ahn SH, *et al.* (2004) Modulation of hepatitis B virus secretion by naturally occurring mutations in the S gene. *J Virol* **78**: 3262–3270.
90. Xu Z, Jensen G, Yen TSB. (1997) Activation of Hepatitis B virus S promoter by the viral large surface protein via induction of stress in the endoplasmic reticulum. *J Virol* **71**: 7387–7392.
91. Surjit M, Jameel S, Lal SK. (2007) Cytoplasmic localization of the ORF2 protein of hepatitis E virus is dependent on its ability to undergo retrotranslocation from the endoplasmic reticulum. *J Virol* **81**: 3339–3345.
92. Schreck R, Rieber P, Baeurele PA. (1991) Reactive oxygen intermediates as apparently widely used messengers in the activation of NFκB transcription factor and HIV. *EMBO J* **10**: 2247–2258.
93. Waris G, Huh KW, Siddiqui A. (2001) Mitochondrially associated hepatitis B virus X protein constitutively activates transcription factors STAT-3 and NFκB via oxidative stress. *Mol Cell Biol* **21**: 7721–7730.
94. Dizdaroglu M. (1994) Chemical determination of oxidative DNA damage by gas chromatography mass spectrometry. *Methods Enzymology* **234**: 3–16.
95. Rahmani Z, Huh KW, Lasher R, Siddiqui A. (2000) Hepatitis B Virus X Protein Colocalizes to Mitochondria with a Human Voltage-Dependent Anion Channel, HVDAC3, and Alters Its Transmembrane Potential. *J Virol* **74**: 2840–2846.
96. Wang JH, Yun C, Kim S, *et al.* (2003) Reactive oxygen species modulates the intracellular level of HBx viral oncoprotein. *Biochem Biophys Res Commun* **310**: 32–39.
97. Tardif KD, Waris G, Siddiqui A. (2005) Hepatitis C virus, ER stress, and oxidative stress. *Trends Microbiol* **13**: 159–163.
98. Moin SM, Pantera M, Jameel S. (2007). The hepatitis E virus (HEV) ORF3 protein protects cells from mitochondrial depolarization and death. *J Biol Chem* (in press).
99. Davis RJ. (2000) Signal transduction to the nucleus by MAP kinase. In J.S. Gutkind (ed), *Signalling Networks and Cell Cycle Control: The Molecular Basis of Cancer and other Diseases*, pp. 153–164. Humana Press, Totowa, New Jersey.

100. Hackel PO, Zwick E, Prenzel N, Ullrich A. (1999) Epidermal growth factor receptors: critical mediators of multiple receptor pathways. *Curr Opin Cell Biol* **11:** 184–189.
101. Miyaki M, Sato C, Sakai K, *et al.* (2000) Malignant transformation and EGFR activation of immortalized mouse liver epithelial cells caused by HBV enhancer-X from a human hepatocellular carcinoma. *Int J Cancer* **85**: 518–522.
102. Benn J, Schneider RJ. (1994) Hepatitis B virus HBx protein activates Ras-GTP complex formation and establishes a Ras, Raf, MAP kinase signaling cascade. *Proc Natl Acad Sci* **91**: 10350–10354.
103. Aoki H, Hayashi, J Moriyama M, *et al.* (2000) Hepatitis C virus core protein interacts with 14-3-3 protein and activates the kinase Raf-1. *J Virol* **74**: 1736–1741.
104. Giambartolomei S, Covone F, Levero M, Balsano C. (2001) Sustained activation of the Raf/MEK/Erk pathway in response to EGF in stable cell lines expressing the hepatitis C virus (HCV) core protein. *Oncogene* **20**: 2606–2610.
105. Koch CA, Anderson D, Moran MF, *et al.* (1991) SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science* **252**: 668–674.
106. Tan SL, Nakao H, He Y, *et al.* (1999) NS5A, a nonstructural protein of hepatitis C virus, binds growth factor receptor-bound protein 2 adaptor protein in a Src homology 3 domain/ligand dependent manner and perturbs mitogenic signaling. *Proc Natl Acad Sci USA* **96**: 5533–5538.
107. He Y, Nakao SL, Polyak SJ, *et al.* (2002) Subversion of cell signaling pathways by hepatitis C virus nonstructural 5A protein via interaction with Grb2 and p85 phosphatidylinositol 3-kinase. *J Virol* **76**: 9207–9217.
108. Francois C, Duverlie G, Rebouillat D, *et al.* (2000) Expression of hepatitis C virus proteins interferes with the antiviral action of interferon independently of PKR-mediated control of protein synthesis. *J Virol* **74**: 5587–5596.
109. Korkaya H, Jameel S, Gupta D, *et al.* (2001) The ORF3 protein of hepatitis E virus binds to src homology 3 domains and activates MAPK. *J Biol Chem* **276**: 42389–42400.
110. Kar-Roy A, Korkaya H, Oberoi R, *et al.* (2004) The hepatitis E virus open reading frame 3 protein activates ERK through binding and inhibition of the MAPK phosphatase. *J Biol Chem* **279**: 28345–28357.

111. Franke TF, Kaplan DR, Cantley LC, Toker A. (1997) Direct regulation of the Akt proto-oncogene product by phosphatidylinositol –3,4-bisphosphate. *Science* **275**: 665–668.
112. Alessi DR, James SR, Downes CP, *et al.* (1997) Characterization of a 3-phosphoinositide-dependent protein kinase which phosphoryates and activates protein kinase B alpha. *Curr Biol* **7**: 261–269.
113. Datta SR, Dudek H, Tao X, *et al.* (1997) Akt phosphorylation of BAD couples survival signals to the cell intrinsic death machinery. *Cell* **91**: 231–241.
114. Cardone MH, Roy N, Stennicke HR, *et al.* (1998) Regulation of cell death protease caspase 9 by phosphorylation. *Science* **282**: 1318–1321.
115. Brunet A, Bonni A, Zigmond MJ, *et al.* (1999) Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**: 857–868.
116. Kops GJ, de Ruiter ND, De Vries-Smits AM, *et al.* (1999) Direct control of the Forkhead transcription factor AFX by protein kinase B. *Nature* **398**: 630–634.
117. Kane LP, Shapiro VS, Stokoe D, Weiss A. (1999) Induction of NFκB by the Akt/PKB kinase. *Curr Biol* **9**: 601–604.
118. Silverman N, Maniatis T. (2001) NFκB signaling pathways in mammalian and insect innate immunity. *Genes Dev* **15**: 2321–2342.
119. Karin M, Lin A. (2002) NF-κB at the cross roads of life and death. *Nature Immunol* **3**: 221–227.
120. Bonte D, Francois C, Castelain S, *et al.* (2004) Positive effect of the hepatitis C virus nonstructural 5A protein on viral multiplication. *Arch Virol* **149**: 1353–1371.
121. Waris G, Livolsi A, Imbert V, *et al.* (2003) Hepatitis C virus NS5A and subgenomic replicon activate NF-kappaB via tyrosine phosphorylation of IkappaBalpha and its degradation by calpain protease. *J Biol Chem* **278**: 40778–40787.
122. You LR, Chen CM, Lee YH. (1999) Hepatitis C virus core protein enhances NF-kB signal pathway triggering by lymphotoxin-β receptor ligand and tumor necrosis factor α. *J Virol* **73**: 1672–1681.
123. Marusawa H, Hijikata M, Chiba T, Shimotohno K. (1999) Hepatitis C virus core protein inhibits Fas- and tumor necrosis factor alpha-mediated apoptosis via NF-kappaB activation. *J Virol* **73**: 4713–4720.
124. Lucas DS, Bartolome J, Amaro MJ, Carreno V. (2003) Hepatitis C virus core protein transactivates the inducible nitric oxide synthase promoter via NF-kappaB activation. *Antiviral Res* **60**: 117–124.

125. Shrivastava A, Manna SK, Ray R, Aggarwal BB. (1998) Ectopic expression of hepatitis C virus core protein differentially regulates nuclear transcription factors. *J Virol* **72**: 9722–9728.
126. Joo M, Hahn YS, Kwon M, *et al.* (2005) Hepatitis C virus core protein suppresses NF-kappaB activation and cyclooxygenase-2 expression by direct interaction with IkappaB kinase beta. *J Virol* **79**: 7648–7657.

Chapter 6

Virus-Cell Interaction of HCV

*Hideki Tani, Yasumasa Komoda, Chang-kwang Limn, Kensuke Suzuki, Kohji Moriishi, Tatsuo Miyamura and Yoshiharu Matsuura**

Hepatitis C virus (HCV) infection is a major health problem worldwide for almost two decades. However, precise mechanisms including an infection of HCV to target cells or a cause of hepatocellular carcinoma are largely unknown. Study of HCV has been hampered by the lack of a robust and reliable cell culture system and a small animal model that supports replication of HCV. Understanding the mechanisms of HCV infection is essential for the development of new effective therapies for chronic hepatitis C. HCV presumably binds to specific receptor(s) and then enters into cells by endocytosis, as do other members of the *Flaviviridae* family and several host membrane proteins have been identified as receptor candidates for HCV. Recent advances of pseudotype virus systems based on vesicular stomatitis virus or retrovirus have provided further information surrounding the initial steps of HCV infection. In this Chapter, we will present current status of our knowledge on the mechanisms of HCV infection including candidate receptors for HCV. (161 words)

Introduction

Hepatitis C virus (HCV) has already infected more than 3% of the worldwide population and 80% of these individuals will develop persistent infection.[17,105] Persistent HCV infection often leads to

*From the Research Institute for Microbial Diseases, Osaka University, Osaka 565-0821, Japan and Department of Virology II, National Institute of Infectious Diseases, Tokyo 162-8640, Japan. E-mail: matsuura@biken.osaka-u.ac.jp

chronic hepatitis, cirrhosis and hepatocellular carcinoma, while acute cases are extremely rare. Although the principal site of HCV replication is thought to be the liver, recent studies revealed that B cell lymphocyte seems to be another site of HCV replication because of the abnormalities of B cells including cryoglobulinemia and an increased risk of non-Hodgkins B-cell lymphoma in hepatitis C patients.[3,90] A comprehensive blood-screening system was established soon after the identification of HCV as a causative agent of non-A, non-B hepatitis. The incidence of post-transfusion hepatitis C has significantly decreased after the introduction of this blood-screening system, however, more than two million people have already been infected with HCV in Japan.

HCV belongs to the family of *Flaviviridae* which includes flaviviruses such as yellow fever virus, dengue virus and West Nile virus and pestiviruses such as bovine viral diarrhea virus and classical swine fever virus. The genome of HCV is comprised of a 9.4 kb single-stranded positive sense RNA that encodes a precursor polyprotein composed of 3010–3030 amino acids. This viral polyprotein is cleaved by a host signal peptidase and viral encoded proteases, resulting in at least 10 viral proteins including capsid (core) protein, two envelope glycoproteins (E1 and E2), p7, nonstructural (NS) protein NS2, NS3, NS4A, NS4B, NS5A and NS5B. The open reading frame of the polyprotein is flanked at both ends by highly conserved untranslated regions (UTR), which are required for viral RNA replication.[32,116] The 5′-UTR harbors an internal ribosome entry site that is essential for Cap independent translation of viral RNA. HCV is grouped into six major genotypes and more than 50 subtypes.[68] Although variability has been documented across the entire genome, the most variable proteins are the envelope glycoproteins. Distinct envelope protein variants are most likely determined at the level of virus-receptor interaction(s).

The lack of an *in vitro* cell culture system has hampered the study of HCV. Based on studies of the assembly and budding of Kunjin virus,[63] it is inferred that HCV particles budded from the endoplasmic reticulum (ER) are released via the secretory pathway. In fact, all of the HCV proteins are associated with the ER or ER-derived membranes, suggesting that genome replication and assembly occur in association with an ER-derived compartment.[27] HCV is thought to interact with a specific cell surface receptor (or receptor complex) for

internalization into the cell by endocytosis. Recent attempts to identify the HCV receptor revealed that several cell surface molecules exhibit binding to HCV envelope proteins as described below.

The putative life cycle of HCV is shown in Fig. 1. At first, HCV particles are trapped by glycosaminoglycans such as heparin or heparan

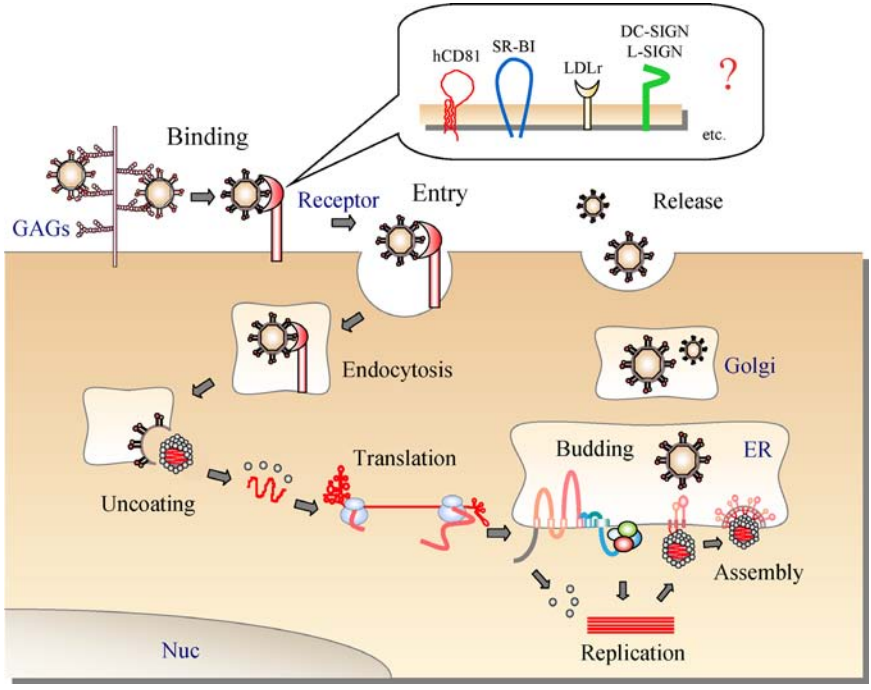


Fig. 1. Putative life cycle of HCV. HCV particles are trapped by glycosaminoglycans (GAGs) on the cell surface, transferred to cell surface receptor and/or co-receptor, and then internalized into cells through endocytosis. hCD81, SR-B1, LDLr, DC-SIGN, and L-SIGN are identified as candidates for HCV receptors. Acidification leads to conformational changes of the HCV envelope proteins into a fusion competent state and induces the fusion of viral membranes with host membranes. Fusion allows the viral genome to be liberated into cytosol. After uncoating, viral RNA is translated into a precursor polyprotein that is processed into each viral protein by cellular and viral proteases, and replication takes place by the viral polymerase complex on the ER membrane. HCV particles seem to bud into the ER lumen after assembly of the nucleocapsid with E1 and E2 proteins and to release from the plasma membrane through the secretory pathway.

sulfate and then transferred to a cell surface receptor and/or co-receptor, and internalized into cells through endocytosis. After uncoating, viral RNA is translated into a precursor polyprotein that is processed into each viral protein by cellular and viral proteases. Viral replication utilizes the viral polymerase complex and takes place on the ER membrane. Core protein was shown to directly bind to the viral sense RNA,[95] and it was suggested to form the nucleocapsid. HCV particles seem to bud into the ER lumen after interaction of the nucleocapsid with E1 and E2 proteins. The small hydrophobic peptide p7 was suggested to form an ion channel,[16,39,77] based on the observation that p7 of pestivirus was shown to be essential for virus budding.[41] NS2 is a component of the NS2-NS3 metalloprotease that autocatalyses cleavage of the NS2-NS3 junction. NS3 contributes to the NS2-NS3 protease and contains a separate serine protease that acts with co-factor NS4A to release the remaining NS proteins. NS3 also contains an RNA helicase and nucleotide triphosphatase activity required for viral replication. NS5A was found to be a highly phosphorylated polypeptide that may be involved in resistance to the antiviral effects of interferon (IFN) alpha.[62,99] NS5A plays an important role in viral replication, because mutation of the phosphorylation sites of NS5A resulted in the enhancement of replication in the HCV replicon cell lines.[62] NS5B has an activity of the RNA dependent RNA polymerase and forms replication complex with other NS proteins and host cellular proteins.[89]

Advances in the treatment of chronic hepatitis C, antiviral agents and their targets were recently described in a review by Walker and colleagues.[111] Here, we will focussed on our knowledge on the mechanisms of HCV infection by using surrogate systems.

Initial Stage of HCV Infection

Host cell surface molecules required for virus entry are classified as either receptor or co-receptor. Several viruses utilize one molecule as a receptor for entry into host cells, while many viruses require co-receptor that localize near the receptor for their entry. Receptors are primarily

involved in the specific attachment of virus to host cells and, in some cases, in viral entry. Affinity of viral envelope proteins with a particular receptor may generally restrict host range, tropism and pathogenicity, although different viruses that have different or partially crossed pathogenicity may utilize the same receptors. Receptors need not to be membrane proteins, as carbohydrates or lipids have been identified as receptors for different viruses.[91] HCV E1 and E2 proteins are the viral component thought to be present at the surface of the viral particles. E1 and E2 proteins contain a large ectodomain region in the N-terminus and a hydrophobic transmembrane region including retention signals in the C-terminus, therefore these proteins are normally retained in the ER membrane.[22] The ectodomains of E1 and E2 proteins are post-translationally modified by extensive N-linked glycosylation.[26,31,109] E1 and E2 proteins have 5 and 11 potential glycosylation sites, respectively, and then these glycans play an important role in the folding of E1 and E2 proteins.[38] Several molecules including human CD81 (hCD81),[80] low-density lipoprotein receptor (LDLr),[2] scavenger receptor class B type I (SR-BI),[88] dendritic cell-specific intracellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN)/liver and lymphnode-specific intracellular adhesion molecule 3-grabbing nonintegrin (L-SIGN),[34] asialoglycoprotein receptor,[87] have been identified to be candidates for a receptor or co-receptor for viral entry. However, the final determination of receptor or co-receptor requirements for HCV is difficult because of the lack of a reliable cell culture systems and a sufficient amount of native viral particles. Three surrogate systems have been developed to study the initial step of HCV infection (Fig. 2). Soluble truncated E2 protein,[30,59,80,81,88] liposomes reconstituted with E1 and E2 proteins,[51] HCV-like particles (HCV-LPs)[106,112] expressed in insect cells and authentic HCV particles in sera of patients have been used to study to identify binding receptors molecules. Cell fusion assay[98] was established to examine the membrane fusion activity of HCV envelope proteins after modifying the original method.[28]

Pseudotype virus systems based on vesicular stomatitis virus (VSV),[50,66] influenza virus,[29] retroviruses,[9] and lentivirus[44] have also been established to identify entry receptors for HCV.

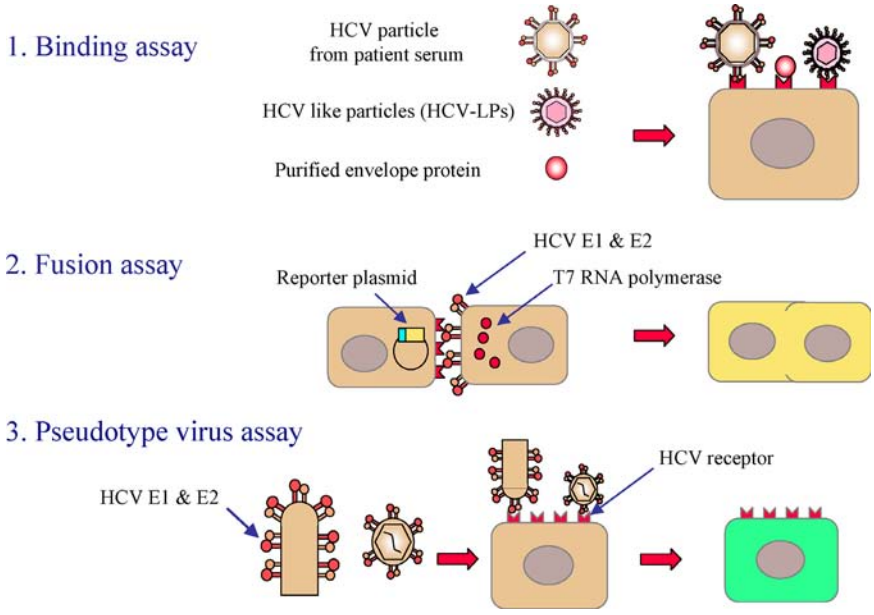


Fig. 2. Surrogate systems to examine infection mechanisms of HCV. There are three systems for the study of infection mechanisms of HCV at the moment. First is a binding assay. Purified envelope protein, HCV-LPs or HCV particles were used as probes to determine the binding receptors. However, this binding assay cannot analyze further steps of infection such as fusion and penetration. Second is a fusion assay. Cell lines expressing HCV envelope proteins and T7 RNA polymerase were co-cultured with the target cell lines containing a reporter plasmid. Upon membrane fusion, T7 RNA polymerase activates the reporter gene. Third is a pseudotype system. Pseudotype viruses based on VSV or retroviruses possessing HCV E1 and E2 proteins can investigate the mechanism of the penetration step by envelope proteins of HCV.

HCV-LPs for HCV-cell Interaction

In the absence of purified HCV particles, virus-like particles (VLPs) have shown to be one of the surrogate models for HCV studies. It has been shown that recombinant expression of the major structural proteins of various viruses leads to the formation of VLPs. Production of VLPs has mostly succeeded by using a recombinant baculovirus expression system. Several viruses are demonstrated the application of

a subunit vaccine,[74,117] the investigation of virus-cell interaction or the identification of candidate receptors by using VLPs produced by baculoviral vectors.[58,100] In the HCV study, several groups have described the generation of HCV-LPs in insect cells using recombinant baculoviruses containing HCV structural genes.[11,19,106,115] HCV-LPs have shown to be effective for stimulating both cellular and humoral immune responses[54,73] and antigenic analyses revealed that HCV-LPs interact with anti-HCV antibodies against nonlinear or conformational epitopes of E1 or E2 proteins,[19,48,96] suggesting that the HCV-LPs seem to resemble immunologically to HCV particles. In addition, studies for interaction between HCV-LPs and host cells revealed that HCV-LPs exhibited dose-dependent and saturable binding to lymphoma and hepatoma cell lines.[106,107,112,115] Although cell surface glycosaminoglycans (GAGs) are involved in the binding of HCV-LPs as recombinant E2 proteins,[6] the binding activity to hCD81 is different between HCV-LPs and the purified E2 proteins,[107] suggesting that conformation of HCV envelope proteins is important for the recognition of the cell surface binding molecules. Therefore, HCV-LPs are useful not only for vaccine candidate but also for a tool to examine the interaction between host cells and HCV.

Pseudotype Virus Systems for Characterization of HCV Infection

A pseudotype virus defines as an enveloped viral particle harboring other type of viral envelope proteins instead of its own envelopes which are not coded into its viral genome.[86,118] A pseudotype virus system based on VSV have been established and applied as characterization of viral entry mechanisms,[97,103] neutralization test for antibodies,[75] vaccine development[83,85] or identification of putative viral receptors[104] for many viruses. The most advantage of the use of this pseudotype virus system is to study the entry mechanisms of many different viruses that are either difficult to amplify in cultured cells or animals or that require high-level containment facilities, i.e. to handle with "biosafety level 4" viruses. In addition, murine leukemia virus (MLV) or human immunodeficiency virus (HIV) have also been engineered to establish pseudotype virus systems.

VSV has a non-segmented 11 kb genome of negative stranded RNA which is transcribed in the cytoplasm and encodes five structural proteins. As reason for many advantages, VSV has been used as a model system for studying the replication and assembly of enveloped RNA viruses. VSV is known to efficiently propagate in many animal cells and readily form pseudotype virions with the envelope proteins from many other different viruses. Reverse genetics systems to recover rabies virus and VSV from cDNA clones have become available.[53,93,113] Recombinant VSV in which native envelope protein G is replaced with foreign viral envelope protein or other membrane protein could contribute to the study of viruses that inefficiently replicate in experimental systems. Additionally, such pseudotype viruses could lead to the induction of cellular and humoral host immunity.[92]

In HCV study, we and others have been adapted to utilize the VSV pseudotype systems,[50,66,101] and the pseudotype viruses possessing modified or unmodified HCV envelope proteins are infectious to human hapatoma cell lines. HCV pseudotype VSVs can be produced by the infection of VSVΔG*-G which consists of ΔG* viruses possessing reporter gene in place of envelope G gene complemented in trans with VSVG protein (Fig. 3). HCV E1 and E2 proteins are normally retained in the ER by C-terminal retention signals.[20,21] In the VSV pseudotype systems, foreign envelope proteins should be expressed on cell surface because VSV normally bud from the plasma membrane. Pseudotype VSVs possessing chimeric proteins comprised of the ectodomains of HCV envelope proteins with the signal sequence, transmembrane and cytoplasmic regions of VSV G envelope protein were constructed.[66,98] The chimeric E1 and E2 proteins were translocated onto the cell surface and incorporated into the released VSV particles. This pseudotype VSVs possessing chimeric E1 and E2 proteins could infect to human hepatoma cell lines HepG2 and Huh7, human kidney cell line 293T, monkey kidney cell lines COS7 and CV-1, but neither mouse, rat nor hamster cell lines. Furthermore, chimeric HCV envelope proteins expressing CHO cell lines induced membrane fusion with HepG2 cells in a pH-dependent manner,[98] confirmed that chimeric E1 and E2 proteins were functionally expressed on the cell surface.

Recently, a recombinant VSV encoding foreign gene instead of VSVG gene has been developed.[33,76] Although lack of infectivity of

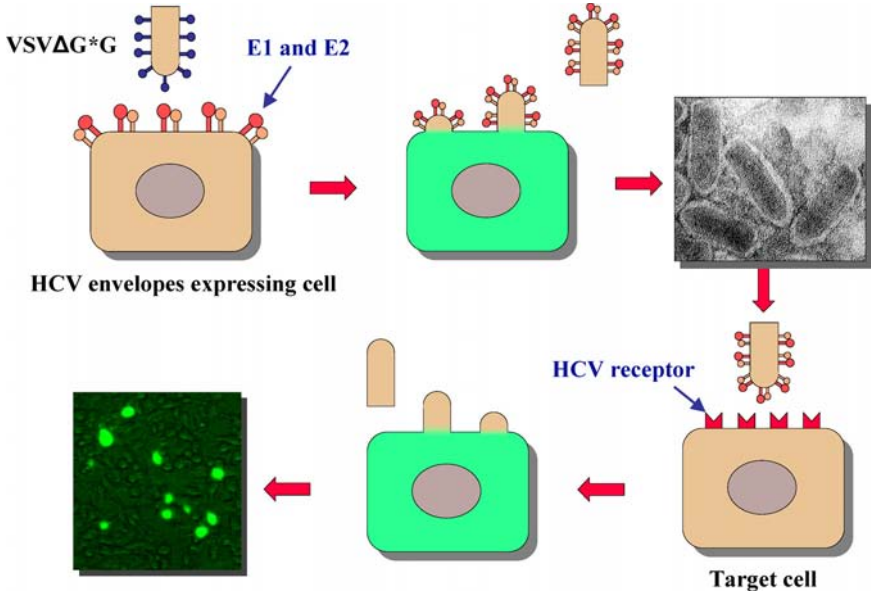


Fig. 3. Construction of pseudotype VSVs. VSVΔG*G, lacking VSVG gene but possessing the VSVG protein on the viral particles, infects to cells expressing E1 and E2 proteins. After removing unbound viruses, infected cells are cultured for one day and pseudotype VSVs are recovered. If target cells had a receptor for the pseudotype viruses, the viruses could infect and express green fluorescent protein (GFP).

recombinant VSVs possessing chimeric HCV envelope proteins to various mammalian cell lines has been reported by Rose and colleagues,[14] molecular size of the chimeric E1 protein incorporated into the recombinant VSVs is smaller than that observed in the pseudotype VSVs. Therefore glycosylation or other post-translational modifications of HCV envelope proteins expressed by different systems or cell lines might determine the infectivity of recombinant VSVs.

Pseudotype particles based on MLV or HIV possessing unmodified HCV envelope proteins have been generated.[9,44] Although HCV envelope proteins are believed to retain in the ER,[20] Bartosch *et al.* and another group reported that the native form of HCV glycoproteins were partially expressed on cell surface of 293T cells and encapsulated retrovirus nucleocapsid, resulting in the production of infectious particles.[9,44] These pseudotype particles produced by transfection of 293T cells with

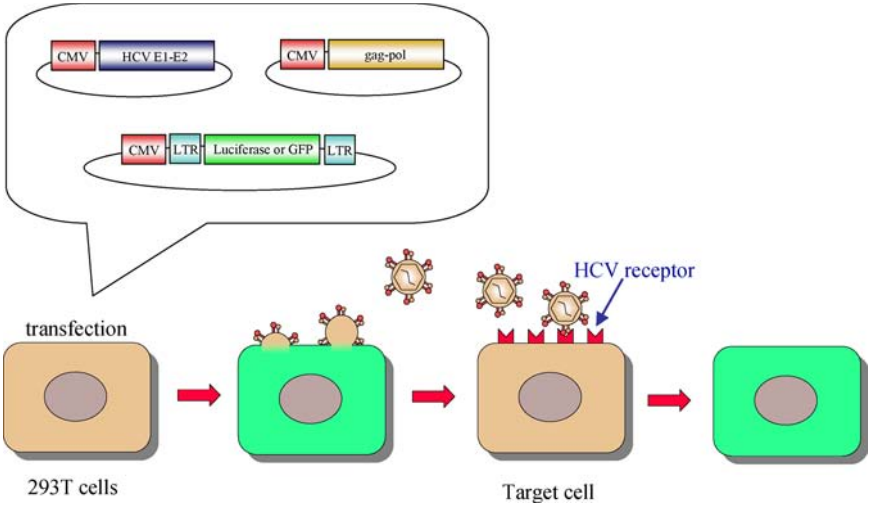


Fig. 4. Construction of pseudotype retroviruses. 293T cells are transfected with three plasmids encoding HCV E1 and E2 proteins, retroviral gag-pol proteins, and a packaging competent GFP, respectively. After 2 to 3 days incubation, pseudotype retroviruses secreted into the culture supernatants. Infectivity of pseudotype viruses containing Gag, Pol, E1 and E2 proteins, as well as the RNA encoding the GFP or luciferase gene is evaluated by the expression of GFP or Luciferase in target cells.

three kinds of expression plasmids (Fig. 4) are infectious to Huh7 cells but not to HepG2 cells. In addition, infection of the pseudotype particles was neutralized by anti-E2 monoclonal antibodies as well as sera of most of patients infected with HCV. Therefore, pseudotype retroviruses also represent a useful model for the study of the early steps of HCV life cycle in combination with the pseudotype VSVs. It was also shown that infection of the pseudotype retroviruses was required both E1 and E2 proteins and no significant difference on the infectivity and neutralization properties was observed among HCV genotypes.[8,9,44,52] Although pseudotype retrovirus infection was inhibited by the addition of anti-hCD81 antibody or a soluble form of hCD81, expression of hCD81 alone failed to confer mouse NIH3T3 cells susceptible to infection by pseudotype MLVs.[10] A similar study was demonstrated that pseudotype HIVs possessing unmodified HCV envelope proteins infect Huh7 cells

in a pH-dependent manner, and, furthermore, the expression of either or all of the candidate receptor molecules, such as hCD81, LDLr, SR-BI and DC-SIGN failed to confer permissivity to pseudotype infection, suggesting that unknown additional molecule(s) are necessary for pseudotype HIVs entry.[44]

Since pseudotype VSVs and retroviruses exhibit a different cell tropism for infection, each pseudotype virus might have different characteristics for the envelope proteins and then utilize different cellular receptors. One possibility is that chimeric E1 protein of the pseudotype VSVs and authentic E1 protein of the pseudotype retroviruses are revealed to be modified into complex-type and high-mannose-type carbohydrates, respectively, suggesting that carbohydrate modification of E1 protein confers cell tropisms of the pseudotype particles. Although a single chain antibody against E2 antibody obtained from a chronic hepatitis C patient exhibited partial neutralization of pseudotype VSVs,[15] neutralization of pseudotype VSVs by mouse monoclonal antibodies against HCV E1 and E2 proteins or patient sera was not observed.[66,101] This is clear contrast to a high prevalence of neutralization antibody against pseudotype retrovirus particles in patients of hepatitis C.[8] One of the characteristics of HCV infection is an establishment of a persistent infection, therefore HCV virions like pseudotype retroviruses might be easily eliminated by the neutralization antibody present in most of hepatitis C patients. Further study to determine the cell tropism of VSV and retrovirus pseudotype particles and the neutralization activity of antibodies raised against E1 and E2 proteins in HCV patients are needed.

1) *Glycosaminoglycans*

Glycosaminoglycans (GAGs) are ubiquitously present on the cell surface, acquiring a net negative charge through N and O sulfation.[13] A variety of viruses[69,91] and other microorganisms[110] have been shown to interact with GAGs, especially heparin or heparan sulfate, in an early step of binding or entry to the host cells. In case of *Flaviviridae* family, flaviviruses such as dengue virus[18,37,43] and Japanese Encephalitis virus[55,56] and pestiviruses such as classical swine fever virus[45] and

bovine viral diarrhea virus[46] are attached and entered to the cells with involvement of GAGs. An early interaction between HCV and cell surface GAGs has also been demonstrated by several experiments using envelope proteins, HCV-LPs or pseudotype viruses. Several groups showed that heparin or heparan sulfate but not other GAGs such as keratan sulfate, chondroitin sulfate, or dermatan sulfate inhibited the cellular binding of recombinant E2 protein and the entry of HCV-LPs[6] or the entry of pseudotype viruses[70,101] in a dose-dependent manner. Moreover, GAG lyases (i.e. heparinase or heparitinase) also inhibited not only E2 binding and HCV-LPs entry[6] but also cell fusion activity[98] and pseudotype virus entry.[66,101] Another experiments revealed that interactions of GAGs with E2 proteins are involved in positively charged E2 hypervariable region 1,[6,12,78] suggesting that E2 protein may easily bind to negatively charged components such as heparin and heparan sulfate. HCV may be concentrated by heparin and heparan sulfate on the cell surface at first, and then transferred to the other binding receptor or entry receptor at the next step.

2) *hCD81*

hCD81, a 25-kD membrane-associated protein which is widely expressed in hematopoietic and epithermal cells, was first reported as a receptor candidate for HCV by Pireli and colleagues.[80] This protein belongs to the tetraspanin superfamily, characterized by forming four transmembrane domains, a short intracellular region, and two extracellular loops.[57] The soluble extracellular domain of HCV E2 protein could bind to the second extracellular loop (EC2) or the large extracellular loop (LEL) of hCD81. EC2/LEL is highly conserved in humans and chimpanzees, which are the only known species permissive to HCV infection. In particular, amino acid 186 of hCD81 is critical for E2 binding and is one of three amino acids that differ from African green monkey CD81, which does not support E2 binding.[42] E2 protein could also bind to tamarin CD81 (the same amino acid at position 186 with hCD81) with higher affinity than hCD81, suggesting that species permissiveness to infection of HCV is not due to CD81. Analyses of the E2 region involved in the CD81 interaction

revealed that the binding site is of a conformational nature involving amino acids 480–493 and 544–551 within the E2 protein.[30,31] Recently, the LEL domain of hCD81 has been crystallized in a hexagonal form at 1.6 Å resolution.[49] hCD81-LEL displays a mushroom-shaped structure (stalk and head subdomains) composed of five a-helices and forms a dimmer in the crystallographic asymmetric unit. This head subdomain exposed to surface region of solvent has the amino acid residues which are essential for the binding of E2 protein. Although the crystal structure of the E2 protein has not been determined yet, crystallography studies might become clearer the interaction of E2 protein and hCD81.

Genotype specificity of affinity of E2 to hCD81 has been reported that E2 protein of H strain exhibited high affinity to hCD81with 10^{-9} M,[79] whereas those cloned from other genotypes showed lower affinity to hCD81.[84,94]

The most important question to be addressed is whether expression of hCD81 alone is sufficient for HCV infection. Replacement of mouse CD81 with hCD81 did not confer mouse susceptible to HCV infection.[65] Although hCD81 can bind to HCV E2 with a *Kd* value of 1.8 nM, it may simply serve as an attachment molecule because hCD81 does not efficiently internalize ligands.[79] HCV E2 protein, HCV-LPs and pseudotype VSVs also bound or entered to the hepatoblastoma cell line HepG2 that does not express detectable levels of hCD81.[66,79,112]

In contrast, pseudotype retroviruses exhibited high susceptibility to the hepatocarcinoma cell line Huh7 which expresses hCD81. The infection of the pseudotype virus was blocked by soluble recombinant hCD81 or by knockdown of hCD81 expression by small interfering RNAs (siRNAs).[119] Although pseudotype retroviruses failed to infect HepG2 cells, the same cells became susceptible by transduction of hCD81.[10,24,67] However, non-hepatic cells engineered to express hCD81 were not permissive and the binding of pseudotype viruses were not correlated to the expression of hCD81, suggesting that hCD81 is one of the important factors involved in the pseudotype retrovirus infection, but other factors present in human hepatocytes are still required for the virus entry. Thus, hCD81 together with additional unknown hepatocyte specific molecule(s), may contribute to the virus-cell binding and/or entry process.

3) *LDLr*

HCV has an unusually low buoyant density in plasma due to an association with various forms of lipoprotein, in particular LDL.[64,71,82] Several groups have recently been reported that LDLr is a cell surface molecule mediating attachment and endocytosis of HCV associated with very-low-density lipoprotein (VLDL) or LDL in serum.[2,5,72] In contrast, HCV purified from plasma fractions without VLDL or LDL did not exhibit measurable LDLr-mediated binding or endocytosis. Apolipoprotein B and apolipoprotein E integrated in LDL or VLDL are essential for binding to the LDLr, because antibody against apolipoprotein B or apolipoprotein E could inhibit the interaction between HCV and LDLr.[2] Interestingly, the interaction between HCV particles from patient plasma and the LDLr was not correlated to the interaction of HCV E2 protein with LDLr.[114] The binding of HCV particles in LDL fractions from patient serum to MOLT-4 cells was higher than those in intermediate-density lipoprotein (IDL) fractions. This binding was correlated with LDLr expression and inhibited by LDL but not soluble hCD81, whereas the binding of E2 protein to MOLT-4 cells was not inhibited by LDL but soluble hCD81 is completely inhibited.[114] Another group was also demonstrated that HCV RNA-containing particles were mostly presented in the LDL fractions among the fractions of VLDL, IDL, and LDL. Binding of these particles was also inhibited by VLDL and LDL or anti-apolipoprotein B and E antibodies, whereas their internalization were increased by the upregulation of LDLr.[5] In contrast, HCV-LPs exhibited binding to human hepatoma and lymphoma cell lines in dose-dependent and saturable manner, but did not correlate with LDLr expression. Although LDLr expression was upregulated in MOLT-4 cells cultured under the lipoprotein-deficient condition, HCV-LPs binding was markedly decreased, suggesting that HCV-LPs bind to hepatoma or lymphoma cell lines independently of LDLr expression.[106,112]

In addition, infection of pseudotype MLVs was not inhibited by VLDL and LDL on Huh7 cells which express LDLr.[9] On the other hand, pseudotype VSVs possessing chimeric E1 protein alone were partially inhibited by excess amount of LDL, suggesting the interaction

of E1 protein with LDLr family.[70] The involvement of the LDLr in HCV infection is still controversial and further studies are needed.

4) *SR-BI*

Scarselli and colleagues reported that SR-BI is an another candidate of receptor for HCV by using E2 protein binding assay.[88] SR-BI is a cell surface membrane protein of CD36 superfamily that binds to chemically modified lipoproteins and often many other types of ligands.[1] HCV E2 proteins of both 1a and 1b genotypes could bind to HepG2 cells which lack hCD81 expression, suggesting that receptor(s) other than hCD81 may participate in binding of E2 protein to HepG2 cells. The E2 proteins could also bind to CHO cells stably expressing human SR-BI, but not mouse SR-BI, through their hyper variable region 1 (HVR1). Expression levels of SR-BI in various cell lines were not correlated with the infectivity of pseudotype MLVs even in the presence of hCD81 and LDLr. Although CHO cell lines expressing SR-BI and/or hCD81 were not permissive to pseudotype MLVs infection, 293T cells over-expressing SR-BI resulted in enhancement of infection of pseudotype MLVs.[10,52] Furthermore, knockdown of SR-BI expression by siRNAs and anti-SR-BI antibodies reduced and inhibited the infection of the pseudotype MLVs, respectively. Recently, it was reported that infection of the pseudotype MLVs was increased by high density lipoproteins (HDLs) but not by lipid-free apoA-I and apoA-II and was suppressed by knockdown of SR-BI or lipid transport inhibitors.[108] Furthermore, HDL-mediated enhancement of pseudotype MLVs infection was not observed in the pseudotype deleted in HVR1 of E2 protein. Further studies are needed to clarify the involvement of human SR-BI in HCV infection.

5) *DC-SIGN and L-SIGN*

Dendritic cells (DCs) are specialized for the recognition of pathogens and have a pivotal role in the control of immunity. Recently, several C-type lectin, DC-SIGN (CD209), and lectin-like receptors, L-SIGN (CD209L), have been characterized that they are expressed abundantly

on the surface of these professional antigen-presenting cells. DC-SIGN and L-SIGN serve as antigen receptors and regulate the migration of DCs and interaction with lymphocytes. DC-SIGN and L-SIGN also serve as capture receptors for several microorganisms such as human cytomegalovirus,[40] Ebola virus,[4,23] HIV-type 1,[35] Dengue virus,[102] *Mycobacterium tuberclosis*,[36] and *Leishmania amastigotes*,[23] through high mannose oligosaccharides of the pathogens.

Several groups have recently suggested that DC-SIGN and L-SIGN are utilized as HCV receptors.[25,34,59,60,81] They demonstrated that recombinant E2 proteins, HCV particles derived from patient serum could specifically bind to cells expressing DC-SIGN or L-SIGN. Although HCV-LPs and pseudotype MLVs also efficiently bound to DCs, preincubation of DCs with C-type lectins, antibodies against DC-SIGN or DC-SIGN could not inhibit HCV-LPs binding or pseudotype MLVs entry, suggesting that C-type lectins were not sufficient to mediate HCV binding and entry.[7,24,44,61,81] In our previous data, pseudotype VSVs could enter myeloid DCs but not plasmacytoid DCs through the interaction with lectin, however, treatment with anti-DC-SIGN antibody did not inhibit the infection of pseudotype VSVs.[47] Therefore, DC-SIGN or L-SIGN may serve as one of the attachment molecules of HCV but it is not clear in the entry process.

Acknowledgments

This work was supported in part by Grants-in-aid from the Ministry of Health, Labor and Welfare, the program for Promotion of Fundamental Studies in Health Sciences of the Organization for Drug ADR Relief, R&D Promotion, and Product Review of Japan, the Ministry of Education, Culture, Sports, Science and Technology, and 21st Century Center of Excellence Program of Japan.

References

1. Acton S, Rigotti A, Landschulz KT, *et al.* (1996) Identification of scavenger receptor SR-BI as a high density lipoprotein receptor. *Science* **271**: 518–520.

2. Agnello V, Abel G, Elfahal M, *et al.* (1999) Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *Proc Natl Acad Sci USA* **96**: 12766–12771.
3. Agnello V, Chung RT, Kaplan LM. (1992) A role for hepatitis C virus infection in type II cryoglobulinemia. *New England Journal of Medicine* **327**: 1490–1495.
4. Alvarez CP, Lasala F, Carrillo J, *et al.* (2002) C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans. *J Virol* **76**: 6841–6844.
5. Andre P, Komurian-Pradel F, Deforges S, *et al.* (2002) Characterization of low- and very-low-density hepatitis C virus RNA-containing particles. *J Virol* **76**: 6919–6928.
6. Barth H, Schäfer C, Adah MI, *et al.* (2003) Cellular binding of hepatitis C virus envelope glycoprotein E2 requires cell surface heparan sulfate. *J Biol Chem* **278**: 41003–41012.
7. Barth H, Ulsenheimer A, Pape GR, *et al.* (2005) Uptake and presentation of hepatitis C virus-like particles by human dendritic cells. *Blood* DOI 10.1182/blood-2004–05 1952 (in press).
8. Bartosch B, Bukh J, Meunier J-C, *et al.* (2003a) *In vitro* assay for neutralizing antibody to hepatitis C virus: evidence for broadly conserved neutralization epitopes. *Proc Natl Acad Sci USA* **100**: 14199–14204.
9. Bartosch B, Dubuisson J, Cosset F-L. (2003b) Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes. *J Exper Med* **197**: 633–642.
10. Bartosch B, Vitelli A, Granier C, *et al.* (2003c) Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. *J Biol Chem* **278**: 41624–41630.
11. Baumert TF, Ito S, Wong DT, Liang TJ. (1998) Hepatitis C virus structural proteins assemble into viruslike particles in insect cells. *J Virol* **72**: 3827–3836.
12. Basu A, Beyene A, Meyer K, Ray R. (2004) The hypervariable region 1 of the E2 glycoprotein of hepatitis C virus binds to glycosaminoglycans, but this binding does not lead to infection in a pseudotype system. *J Virol* **78**: 4478–4486.
13. Bernfield M, Gotte M, Park PW, *et al.* (1999) Functions of cell surface heparan sulfate proteoglycans. *Ann Rev Biochem* **68**: 729–777.
14. Buonocore L, Blight KJ, Rice CM, Rose JK. (2002) Characterization of vesicular stomatitis virus recombinants that express and incorporate high levels of hepatitis C virus glycoproteins. *J Virol* **76**: 6865–6872.

15. Burioni R, Matsuura Y, Mancini N, *et al.* (2002) Diverging effects of human recombinant anti-hepatitis C virus (HCV) antibody fragments derived from a single patient on the infectivity of a vesicular stomatitis virus/HCV pseudotype. *J Virol* **76**: 11775–11779.
16. Carrere-Kremer S, Montpellier-Pala C, Cocquerel L, *et al.* (2002) Subcellular localization and topology of the p7 polypeptide of hepatitis C virus. *J Virol* **76**: 3720–3730.
17. Cerny A, Chisari FV. (1999) Pathogenesis of chronic hepatitis C: immunological features of hepatic injury and viral persistence. *Hepatology* **30**: 595–601.
18. Chen Y, Maguire T, Hileman RE, *et al.* (1997) Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat Med* **3**: 866–871.
19. Clayton RF, Owsianka A, Aitken J, *et al.* (2002) Analysis of antigenicity and topology of E2 glycoprotein present on recombinant hepatitis C virus-like particles. *J Virol* **76**: 7672–7682.
20. Cocquerel L, Duvet S, Meunier JC, *et al.* (1999) The transmembrane domain of hepatitis C virus glycoprotein E1 is a signal for static retention in the endoplasmic reticulum. *J Virol* **73**: 2641–2649.
21. Cocquerel L, Meunier JC, Pillez A, *et al.* (1998) A retention signal necessary and sufficient for endoplasmic reticulum localization maps to the transmembrane domain of hepatitis C virus glycoprotein E2. *J Virol* **72**: 2183–2191.
22. Cocquerel L, Op de Beeck A, Lambot M, *et al.* (2002) Topological changes in the transmembrane domains of hepatitis C virus envelope glycoproteins. *EMBO J* **21**: 2893–2902.
23. Colmenares M, Puig-Kroger A, Pello OM, *et al.* (2002) Dendritic cell (DC)-specific intercellular adhesion molecule 3 (ICAM-3)-grabbing nonintegrin (DC-SIGN, CD209) a C-type surface lectin in human DCs, is a receptor for Leishmania amastigotes. *J Biol Chem* **277**: 36766–36769.
24. Cormier EG, Durso RJ, Tsamis F, *et al.* (2004) L-SIGN (CD209L) and DC-SIGN (CD209) mediate transinfection of liver cells by hepatitis C virus. *Proc Natl Acad Sci USA* **101**: 14067–14072.
25. Cormier EG, Tsamis F, Kajumo F, *et al.* (2004) CD81 is an entry coreceptor for hepatitis C virus. *Proc Natl Acad Sci USA* **101**: 7270–7274.
26. Dubuisson J. (2000) Folding, assembly and subcellular localization of HCV glycoproteins. *Curr Topics Microbiol Immunol* **242**: 135–148.

27. Dubuisson J, Penin F, Moradpour D. (2002) Interaction of hepatitis C virus proteins with host cell membranes and lipids. *Trends Cell Biol* **12**: 517–523.
28. Feng Y, Broder CC, Kennedy PE, Berger EA. (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**: 872–877.
29. Flint M, Thomas JM, Maidens CM, *et al.* (1999a) Functional analysis of cell surface-expressed hepatitis C virus E2 glycoprotein. *J Virol* **73**: 6782–6790.
30. Flint M, Maidens C, Loomis-Price LD, *et al.* (1999b) Characterization of hepatitis C virus E2 glycoprotein interaction with a putative cellular receptor, CD81. *J Virol* **73**: 6235–6244.
31. Flint M, McKeating JA. (2000) The role of the hepatitis C virus glycoproteins in infection. *Rev Med Virol* **10**: 101–117.
32. Friebe P, Lohmann V, Krieger N, Bartenschlager R. (2001) Sequences in the 5′ nontranslated region of hepatitis C virus required for RNA replication. *J Virol* **75**: 12047–12057.
33. Garbutt M, Liebscher R, Wahl-Jensen V, *et al.* (2004) Properties of replication-competent vesicular stomatitis virus vectors expressing glycoproteins of filoviruses and arenaviruses. *J Virol* **78**: 5458–5465.
34. Gardner JP, Durso RJ, Arrigale RR, *et al.* (2003) L-SIGN (CD 209L) is a liver-specific capture receptor for hepatitis C virus. *Proc Natl Acad Sci USA* **100**: 4498–4503.
35. Geijtenbeek TB, Kwon DS, Torensma R, *et al.* (2000) DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100**: 587–597.
36. Geijtenbeek TB, Van Vliet SJ, Koppel EA, *et al.* (2003) Mycobacteria target DC-SIGN to suppress dendritic cell function. *J Exper Med* **197**: 7–17.
37. Germi R, Crance J-M, Garin D, *et al.* (2002) Heparan sulfate-mediated binding of infectious dengue virus type 2 and yellow fever virus. *Virology* **292**: 162–168.
38. Goffard A, and Dubuisson J. (2003) Glycosylation of hepatitis C virus envelope proteins. *Biochimie* **85**: 295–301.
39. Griffin SD, Beales LP, Clarke DS, *et al.* (2003) The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. *FEBS Lett* **535**: 34–38.
40. Halary F, Amara A, Lortat-Jacob H, *et al.* (2002) Human cytomegalovirus binding to DC-SIGN is required for dendritic cell infection and target cell trans-infection. *Immunity* **17**: 653–664.

41. Harada T, Tautz N, Thiel HJ. (2000) E2-p7 region of the bovine viral diarrhea virus polyprotein: processing and functional studies. *J Virol* **74**: 9498–9506.
42. Higginbottom A, Quinn ER, Kuo CC, *et al.* (2000) Identification of amino acid residues in CD81 critical for interaction with hepatitis C virus envelope glycoprotein E2. *J Virol* **74**: 3642–3649.
43. Hilgard P, Stockert R. (2000) Heparan sulfate proteoglycans initiate dengue virus infection of hepatocytes. *Hepatology* **32**: 1069–1077.
44. Hsu M, Zhang J, Flint M, *et al.* (2003) Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles. *Proc Natl Acad Sci USA* **100**: 7271–7276.
45. Hulst MM, van Gennip HGP, Vlot AC, *et al.* (2001) Interaction of classical swine fever virus with membrane-associated heparan sulfate: role for virus replication *in vivo* and virulence. *J Virol* **75**: 9585–9595.
46. Iqbal M, Flick-Smith H, McCauley JW. (2000) Interaction of bovine viral diarrhoea virus glycoprotein E^{rns} with cell surface glycosaminoglycans. *J Gen Virol* **81**: 451–459.
47. Kaimori A, Kanto T, Limn CK, *et al.* (2004) Pseudotype hepatitis C virus enters immature myeloid dendritic cells through the interaction with lectin. *Virology* **324**: 74–83.
48. Keck Z-Y, Sung VMH, Perkins S, *et al.* (2004) Human monoclonal antibody to hepatitis C virus E1 glycoprotein that blocks virus attachment and viral infectivity. *J Virol* **78**: 7257–7263.
49. Kitadokoro K, Bordo D, Galli G, *et al.* (2001) CD81 extracellular domain 3D structure: insight into the tetraspanin superfamily structural motifs. *EMBO J* **20**: 12–18.
50. Lagging LM, Meyer K, Owens RJ, Ray R. (1998) Functional role of hepatitis C virus chimeric glycoproteins in the infectivity of pseudotyped virus. *J Virol* **72**: 3539–3546.
51. Lambot M, Frétier S, Op De Beeck A, *et al.* (2002) Reconstitution of hepatitis C virus envelope glycoproteins into liposomes as a surrogate model to study virus attachment. *J Biol Chem* **277**: 20625–20630.
52. Lavillette D, Tarr AW, Voisset C, *et al.* (2005) Characterization of host-range and cell entry properties of the major genotypes and subtypes of hepatitis C virus. *Hepatology* **41**: 265–274.
53. Lawson ND, Stillman EA, Whitt MA, Rose JK. (1995) Recombinant vesicular stomatitis viruses from DNA. *Proc Natl Acad Sci USA* **92**: 4477–4481.

54. Lechmann M, Satoi J, Vergalla J, *et al.* (2001) Hepatitis C virus-like particles induce virus-specific humoral and cellular immune responses in mice. *Hepatology* **34**: 417–423.
55. Lee E, Hall RA, Lobigs M. (2004) Common E protein determinants for attenuation of glycosaminoglycan-binding variants of Japanese encephalitis and West Nile viruses. *J Virol* **78**: 8271–8280.
56. Lee E, Lobigs M. (2002) Mechanism of virulence attenuation of glycosaminoglycan-binding variants of Japanese encephalitis virus and Murray Valley encephalitis virus. *J Virol* **76**: 4901–4911.
57. Levy S, Todd SC, Maecker HT. (1998) CD81 (TAPA-1): a molecule involved in signal transduction and cell adhesion in the immune system. *Annu Rev Immunol* **16**: 89–109.
58. Li TC, Yamakawa Y, Suzuki K, *et al.* (1997) Expression and self-assembly of empty virus-like particles of hepatitis E virus. *J Virol* **71**: 7207–7213.
59. Lozach P-Y, Lortat-Jacob H, de Lacroix de Lavalette A, *et al.* (2003) DC-SIGN and L-SIGN are high affinity binding receptors for hepatitis C virus glycoprotein E2. *J Biol Chem* **278**: 20358–20366.
60. Lozach P-Y, Amara A, Bartosch B, *et al.* (2004) C-type lectins L-SIGN and DC-SIGN capture and transmit infectious hepatitis C virus pseudotype particles. *J Biol Chem* **279**: 32035–32045.
61. Ludwig IS, Lekkerkerker AN, Depla E, *et al.* (2004) Hepatitis C virus targets DC-SIGN and L-SIGN to escape lysosomal degradation. *J Virol* **78**: 8322–8332.
62. Macdonald A, Harris M. (2004) Hepatitis C virus NS5A: tales of a promiscuous protein. *J Gen Virol* **85**: 2485–2502.
63. Mackenzie JM, Westaway EG. (2001) Assembly and maturation of the flavivirus Kunjin virus appear to occur in the rough endoplasmic reticulum and along the secretory pathway, respectively. *J Virol* **75**: 10787–10799.
64. Manzini P, Dusheiko GM, Brown DJC, *et al.* (1998) Hepatitis C virus tends to associate preferentially with high density lipoproteins by standard ultracentrifugal fractionation of plasma from patients with chronic HCV infection. *Hepatol Res* **11**: 158–165.
65. Masciopinto F, Freer G, Burgio VL, *et al.* (2002) Expression of human CD81 in transgenic mice does not confer susceptibility to hepatitis C virus infection. *Virology* **304**: 187–196.
66. Matsuura Y, Tani H, Suzuki K, *et al.* (2001) Characterization of pseudotype VSV possessing HCV envelope proteins. *Virology* **286**: 263–275.

67. McKeating JA, Zhang LQ, Logvinoff C, *et al.* (2004) Diverse hepatitis C virus glycoproteins mediate viral infection in a CD81-dependent manner. *J Virol* **78**: 8496–8505.
68. Mellor J, Holmes EC, Jarvis LM, *et al.* (1995) Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. *J Gen Virol* **76**: 2493–2507.
69. Mettenleiter TC. (2002) Brief overview on cellular virus receptors. *Virus Res* **82**: 3–8.
70. Meyer K, Beyene A, Bowlin TL, *et al.* (2004) Coexpression of hepatitis C virus E1 and E2 chimeric envelope glycoproteins displays separable ligand sensitivity and increases pseudotype infectious titer. *J Virol* **78**: 12838–12847.
71. Miyamoto H, Okamoto H, Sato K, *et al.* (1992) Extraordinarily low density of hepatitis C virus estimated by sucrose density gradient centrifugation and the polymerase chain reaction. *J Gen Virol* **73**: 715–718.
72. Monazahian M, Bohme I, Bonk S, *et al.* (1999) Low density lipoprotein receptor as a candidate receptor for hepatitis C virus. *J Med Virol* **57**: 223–229.
73. Murata K, Lechmann M, Qiao M, *et al.* (2003) Immunization with hepatitis C virus-like particles protects mice from recombinant hepatitis C virus-vaccinia infection. *Proc Natl Acad Sci USA* **100**: 6753–6758.
74. Noad R, Roy P. (2003) Virus-like particles as immunogens. *Trends Microbiol* **11**: 438–444.
75. Ogino M, Ebihara H, Lee B-H, *et al.* (2003) Use of vesicular stomatitis virus pseudotypes bearing hantaan or seoul virus envelope proteins in a rapid and safe neutralization test. *Clini Diagnost Lab Immunol* **10**: 154–160.
76. Okuma K, Dalton KP, Buonocore L, *et al.* (2003) Development of a novel surrogate virus for human T-cell leukemia virus type 1: inhibition of infection by osteoprotegerin. *J Virol* **77**: 8562–8569.
77. Pavlovic D, Neville DC, Argaud O, *et al.* (2003) The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. *Proc Natl Acad Sci USA* **100**: 6104–6108.
78. Penin F, Combet C, Germanidis G, *et al.* (2001) Conservation of the conformation and positive charges of hepatitis C virus E2 envelope

glycoprotein hypervariable region 1 points to a role in cell attachment. *J Virol* **75**: 5703–5710.
79. Petracca R, Falugi F, Galli G, *et al.* (2000) Structure-function analysis of hepatitis C virus envelope-CD81 binding. *J Virol* **74**: 4824–4830.
80. Pileri P, Uematsu Y, Campagnoli S, *et al.* (1998) Binding of hepatitis C virus to CD81. *Science* **282**: 938–941.
81. Pohlmann S, Zhang J, Baribaud F, *et al.* (2003) Hepatitis C virus glycoproteins interact with DC-SIGN and DC-SIGNR. *J Virol* **77**: 4070–4080.
82. Prince AM, Huima-Byron T, Parker TS, Levine DM (1996). Visualization of hepatitis C virions and putative interfering particles isolated from low-density lipoproteins. *J Viral Hep* **3**: 11–17.
83. Roberts A, Buonocore L, Prince R, *et al.* (1999) Attenuated vesicular stomatitis viruses as vaccine vectors. *J Virol* **73**: 3723–3732.
84. Roccasecca R, Ansuini H, Vitelli A, *et al.* (2003) Binding of the hepatitis C virus E2 glycoprotein to CD81 is strain specific and is modulated by a complex interplay between hypervariable regions 1 and 2. *J Virol* **77**: 1856–1867.
85. Rose NF, Marx PA, Luckay A, *et al.* (2001) An effective AIDS vaccine based on live attenuated vesicular stomatitis virus recombinants. *Cell* **106**: 539–549.
86. Rubin H. (1965) Genetic control of cellular susceptibility to pseudotypes of Rous sarcoma virus. *Virology* **26**: 270–278.
87. Saunier B, Triyatni M, Ulianich L, *et al.* (2003) Role of the asialoglycoprotein receptor in binding and entry of hepatitis C virus structural proteins in cultured human hepatocytes. *J Virol* **77**: 546–559.
88. Scarselli E, Ansuini H, Cerino R, *et al.* (2002) The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J* **21**: 5017–5025.
89. Schmidt-Mende J, Bieck E, Hugle T, *et al.* (2001) Determinants for membrane association of the hepatitis C virus RNA-dependent RNA polymerase. *J Biol Chem* **276**: 44052–44063.
90. Schmidt WN, Stapleton JT, LaBrecque DR, *et al.* (2000) Hepatitis C virus (HCV) infection and cryoglobulinemia: analysis of whole blood and plasma HCV-RNA concentrations and correlation with liver histology. *Hepatology* **31**: 737–744.
91. Schneider-Schaulies J. (2000) Cellular receptors for viruses: links to tropism and pathogenesis. *J Gen Virol* **81**: 1413–1429.

92. Schnell MJ, Buonocore L, Kretzschmar E, *et al.* (1996) Foreign glycoproteins expressed from recombinant vesicular stomatitis viruses are incorporated efficiently into virus particles. *Proc Natl Acad Sci USA* **93**: 11359–11365.
93. Schnell MJ, Mebatsion T, Conzelmann KK. (1994) Infectious rabies viruses from cloned cDNA. *EMBO J* **13**: 4195–4203.
94. Shaw ML, McLauchlan J, Mills PR, *et al.* (2003) Characterisation of the differences between hepatitis C virus genotype 3 and 1 glycoproteins. *J Med Virol* **70**: 361–372.
95. Shimoike T, Mimori S, Tani H, *et al.* (1999) Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation. *J Virol* **73**: 9718–9725.
96. Steinmann D, Barth H, Gissler B, *et al.* (2004) Inhibition of hepatitis C virus-like particle binding to target cells by antiviral antibodies in acute and chronic hepatitis C. *J Virol* **78**: 9030–9040.
97. Takada A, Robison C, Goto H, *et al.* (1997) A system for functional analysis of Ebola virus glycoprotein. *Proc Natl Acad Sci USA* **94**: 14764–14769.
98. Takikawa S, Ishii K, Aizaki H, *et al.* (2000) Cell fusion activity of hepatitis C virus envelope proteins. *J Virol* **74**: 5066–5074.
99. Tan S-L, Katze MG. (2001) How hepatitis C virus counteracts the interferon response: the jury is still out on NS5A. *Virology* **284**: 1–12.
100. Tamura M, Natori K, Kobayashi M, *et al.* (2000) Interaction of recombinant norwalk virus particles with the 105-kilodalton cellular binding protein, a candidate receptor molecule for virus attachment. *J Virol* **74**: 11589–11597.
101. Tamura K, Oue A, Tanaka A, *et al.* (2005) Efficient formation of vesicular stomatitis virus pseudotypes bearing the native forms of hepatitis C virus envelope proteins detected after sonication. *Microbes Infect* **7**: 29–40.
102. Tassaneetrithep B, Burgess TH, Granelli-Piperno A, *et al.* (2003) DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. *J Exper Med* **197**: 823–829.
103. Tatsuo H, Okuma K, Tanaka K, *et al.* (2000a) Virus entry is a major determinant of cell tropism of Edmonston and wild-type strains of measles virus as revealed by vesicular stomatitis virus pseudotypes bearing their envelope proteins. *J Virol* **74**: 4139–4145.

104. Tatsuo H, Ono N, Tanaka K, Yanagi Y. (2000b) SLAM (CDw150) is a cellular receptor for measles virus. *Nature* **406**: 893–897.
105. Theodore D, Fried MW. (2000) Natural history and disease manifestations of hepatitis C infection. *Cur Topics Microbiol Immuno* **242**: 43–54.
106. Triyatni M, Saunier B, Maruvada P, *et al.* (2002) Interaction of hepatitis C virus-like particles and cells: a model system for studying viral binding and entry. *J Virol* **76**: 9335–9344.
107. Triyatni M, Vergalla J, Davis AR, *et al.* (2000b) Structural features of envelope proteins on hepatitis C virus-like particles as determined by anti-envelope monoclonal antibodies and CD81 binding. *Virology* **298**: 124–132.
108. Voisset C, Callens N, Blanchard E, *et al.* (2005) High density lipoproteins facilitate hepatitis C virus entry through the scavenger receptor class B type I. *J Biol Chem* **280**: 7793–7799.
109. Voisset C, Dubuisson J. (2004) Functional hepatitis C virus envelope glycoproteins. *Biol Cell* **96**: 413–420.
110. Wadstrom T, Ljungh A. (1999) Glycosaminoglycan-binding microbial proteins in tissue adhesion and invasion: key events in microbial pathogenicity. *J Med Microbiol* **48**: 223–233.
111. Walker MP, Appleby TC, Zhong W, *et al.* (2003) Hepatitis C virus therapies: current treatments, targets and future perspectives. *Antiviral Chem Chemother* **14**: 1–21.
112. Wellnitz S, Klumpp B, Barth H, *et al.* (2002) Binding of hepatitis C virus-like particles derived from infectious clone H77C to defined human cell lines. *J Virol* **76**: 1181–1193.
113. Whelan SP, Ball JN, Wertz GT. (1995) Efficient recovery of infectious vesicular stomatitis virus entirely from cDNA clones. *Proc Natl Acad Sci USA* **92**: 8388–8392.
114. Wunschmann S, Medh JD, Klinzmann D, *et al.* (2000) Characterization of hepatitis C virus (HCV) and HCV E2 interactions with CD81 and the low-density lipoprotein receptor. *J Virol* **74**: 10055–10062.
115. Xiang J, Wunschmann S, George SL, *et al.* (2002) Recombinant hepatitis C virus-like particles expressed by baculovirus: Utility in cell-binding and antibody detection assays. *J Med Virol* **68**: 537–543.
116. Yanagi M, St Claire M, Emerson SU, *et al.* (1999) *In vivo* analysis of the 3′ untranslated region of the hepatitis C virus after *in vitro* mutagenesis of an infectious cDNA clone. *Proc Natl Acad Sci USA* **96**: 2291–2295.

117. Yao Q, Bu Z, Vzorov A, *et al.* (2003) Virus-like particle and DNA-based candidate AIDS vaccines. *Vaccine* **21**: 638–643.
118. Zavada J. (1976) Viral pseudotypes and phenotypic mixing. *Arch Virol* **50**: 1–15.
119. Zhang J, Randall G, Higginbottom A, *et al.* (2004) CD81 is required for hepatitis C virus glycoprotein-mediated viral infection. *J Virol* **78**: 1448–1455.

Chapter 7

RNA Replication of Hepatitis C Virus

Hideki Aizaki† *and Tetsuro Suzuki**,†

A precursor polyprotein of hepatitis C virus (HCV) is processed by host signalases and viral proteases into at least ten matured proteins. Nonstructural proteins, the NS2-3 protease, the NS3 serine protease, the NS3 helicase, and the NS5B RNA-dependent RNA polymerase function in the polyprotein processing and viral replication. Although studies on HCV replication have been hampered by the absence of a productive cell culture system, establishment of cell cultures that are highly permissive for replication of subgenomic and full-length HCV RNAs has opened new avenues to study the initiation of the viral replication. The use of the culture system has recently generated considerable information on HCV RNA replication, and as a result the formation of an intracellular membrane-associated replication complex (RC) composed of viral and cellular proteins as well as replicating RNA is proposed. Intensive research is being carried out to understand the cellular pathways for HCV RC formation and its biochemical function.

Introduction

Hepatitis C virus (HCV) is the most important causative agent of posttransfusion and sporadic non-A, non-B hepatitis. It is a positive-stranded RNA virus belonging to *Flaviviridae*, genus *Hepacivirus*, approximately 9.6 kb in length. Persistent infection with HCV is

*Corresponding author.
†Department of Virology II, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan.

associated with development of chronic hepatitis, hepatic steatosis, cirrhosis, and hepatocellular carcinoma. There are currently 170 million or more HCV carriers worldwide. In past years, anti-hepatitis C therapy has modestly improved; however, a currently available combination therapy, consisting of interferon (IFN)-α and the nucleoside analogue ribavirin, shows a sustained response in only less than half of the treated patients. The development of innovative treatment alternatives for patients infected with HCV is urgently required, and a better understanding of the mechanism of HCV replication should allow the identification of novel targets for antiviral intervention specific to HCV.

Cell Culture Systems for HCV Replication

Studies on HCV replication have long been hampered by the lack of efficient cell culture systems. Although many attempts have been made to achieve full replication of HCV in cell cultures, all systems including cell lines derived from human hepatoma, B cells and T cells, primary hepatocytes and peripheral blood mononuclear cells suffer from low virus yield and are not robust enough to allow genetic analyses of the HCV life cycle. The development of HCV subgenomic replicons (self-replicating RNAs) has allowed examination of viral RNA replication in cell culture.[1] Expression systems of heterologous virus genes based on RNA replicons have been established in a variety of positive-strand RNA viruses such as polio virus,[2–5] alphaviruses semliki forest virus,[6] sindbis virus,[7–10] kunjin virus,[11] human rhinovirus 14,[12] and bovine viral diarrhea virus.[13] In general, advantages of replicon systems are based on (1) a high level of gene expression and RNA replication, (2) easy construction of recombinants, and (3) the permitting of a wide host range.

The HCV replicons are typically composed of the 5′-nontranslated region (NTR), which directs translation of the gene encoding the neomycin phosohotransferase (Neor); the internal ribosome entry site (IRES) of the encephalomyocarditis virus (EMCV), which directs translation of HCV NS3 through NS5B region; and the 3′-NTR. Following transfection of RNA transcribed from the above bicistronic constructs into a human hepatoma cell line Huh-7, antibiotic G418-resistent cells could be obtained in which the subgenomic RNA replicated autonomously. RNA replication was at first detected at low

frequency, followed by the identification of replicons harboring cell culture-adaptive mutations yielding higher replication efficiency.[14–16] Some of the most adaptive mutations are located at highly conserved serine residues within NS5A upstream of the region putatively involved in IFN sensitivity. A combination of adaptive mutation in NS3 and NS5A resulted in the highest level of replication of particular HCV genotype 1b isolate.[16] Later work, however, has indicated that adaptive mutations can arise in most of the viral nonstructural proteins.[17,18] The mechanisms by which adaptive mutations increase RNA replication efficiency are not well understood.

Huh-7 cells have been permissive for adapted HCV replicons, although variability in the permissiveness for replicons has been observed for these cells. Recently, some studies showed that HCV subgenomic RNAs could replicate in human hepatic HepG2 and IMY-N9 cells, human cervical carcinoma HeLa cells, human embryonic kidney 293 cells,[19,20] a human T cell line MT-2C[21] and mouse hepatoma cells.[22] These findings suggest that host factors required for HCV RNA replication are not hepatocyte specific and are not restricted to cells of human origin.

Finally, taking advantage of cell culture-adaptive mutations, full-length dicistrinic HCV genomes that efficiently and stably replicate in Huh-7 cells have been developed. By immuno-electron microscopy, the core protein expressed in the cells harboring the genomic HCV RNA was localized mainly to the boundary of cytoplasmic lipid storage vesicles and was also found at the endoplastic recticulum (ER).[23] In the cells, viral envelope proteins E1 and E2 formed heterodimers and existed at the ER and cis-Golgi compartments. Cell culture systems based on the selectable subgenomic and genome-length dicistronic HCV RNAs, which produce abundant viral RNA and nonstructural proteins, open avenues of biochemical and genetic studies for HCV replication.

Polyprotein Translation and Processing

The HCV genome carries a 5′-NTR, an open reading frame (ORF) that encodes a polyprotein with a length of ~3010 amino acids, and a 3′-NTR. The precursor polyprotein is co-translationally or post-processed by both viral and host proteases into at least ten viral products.

These viral products include i) structural proteins consisting of the following virus particles: core, E1, and E2, which are encoded at the N terminus; and ii) nonstructural proteins involved in viral RNA replication, which encompass the remainder of the viral polyprotein and consist of NS2, NS3, NS4A, NS4B, NS5A, and NS5B.

The ~340 nucleotide long 5′ NTR is functionally characterized as IRES to direct cap-independent translation of the genome. The HCV 5′ NTR is highly conserved in the genome and the secondary structural models reveal four distinct RNA domains in the region.[24] The IRES has been shown to cover most of the 5′-NTR, with the 5′-end being located in between (nt) 28 and 69. By analogy with other RNA viruses with IRES-mediated expression, the HCV 5′-NTR has been expected to contain not only determinants for translation, but also cis-acting elements for RNA replication. Recent studies demonstrated that i) the sequence upstream of the IRES is essential for the viral RNA replication, ii) sequences within the IRES are required for high-level HCV replication, and iii) the stem-loop domain II of the IRES is crucial for the replication.[25]

The 3′-NTR also contains significant predicted RNA structure with three distinct domains: a variable region (VR) of ~40 nucleotides, a variable length poly(U/UC) tract, and a highly conserved 98-nt 3′ terminal segment (3′X) that putatively forms three stem-loop structures.[26–28] The viral RNA replication was not detected when any of the three putative stem-loop structures within the 3′X region or whole of poly(U/UC) were deleted.[29] The VR segment also contributes to efficient RNA replication.[30]

Nonstructural proteins of HCV are processed by two viral proteases: the junction of NS2 and NS3 is cleaved by the NS2-3 protease, which spans NS2 and the N-terminal domain of NS3; the remaining four junctions are cleaved by the serine protease located at the N-terminal 180 residues of NS3 protein. NS2 protein is highly hydrophobic and the autoproteolytic reaction presumably occurs in a membrane compartment. Deletion of NS2 from the HCV nonstructural polyprotein did not abolish the replication of HCV RNA in cell culture, indicating that it is not required for viral RNA replication.[1,14] NS3 is a multifunctional molecule. Besides its N-terminal protease

activity, the helicase and nucleotide triphosphatase (NTPase) activities reside in the C-terminal 500 residues of the NS3 protein.[31–35] Helicase/NTPase activities are essential for the replication of the HCV genome. The helicase activity is presumed to be involved in unwinding a putative double-stranded replication intermediate or to remove regions of secondary structure so that the RNA-dependent RNA polymerase of NS5B (see below) can copy both strands of the viral RNA. It is likely that the NTPase activity is coupled with the helicase function, supplying the energy required for disrupting RNA duplexes.

NS4A functions as a cofactor of the NS3 serine protease and is required for efficient polyprotein processing. There are significant differences in the stability and activity of the NS3 protease in the presence and absence of NS4A. NS3 protein is relatively unstable when expressed in cells in the absence of NS4A.[36] Structural studies by NMR and X-ray methods indicate that the NS3-4A complex has a more highly ordered N-terminal domain, and NS4A binding leads the NS3 protease to a rearrangement of the active site triad to a canonical conformation.[37] It has been predicted that the N-terminus of NS4A forms a transmembrane helix, which presumably anchors the NS3-4A complex to the cellular membrane.[38] NS4B is a hydrophobic, membrane-associated protein, which colocalizes predominantly with ER markers.[39–42] Although relatively little is known to date about functions of NS4B in the HCV life cycle, inhibitory activities in the translation[43,44] and modulation of NS5B enzymatic function[45] have been reported. It has also been shown that NS4B protein can induce a membranous web consisting of small vesicles embedded in a membranous matrix,[46] and that the newly synthesized HCV RNA exists in these membrane webs and speckles.[47,48] NS4B may play an important role in the formation of the HCV RNA replication complex.[49]

NS5A is a phosphoprotein, which is mainly phophorylated on serine residues[50] mediated by one or more cellular serine-threonine kinases.[51–54] The role of NS5A phosphorylation in HCV replication is so far not clear; however, it has been thought that the protein is important for the viral lifecycle since phosphorylation of NS5A is a

conserved feature among different HCV isolates and among other members of the *Flaviviridae*. Evidence indicating an involvement of NS5A in the viral RNA replication is now accumulating. A hot spot of the cell culture-adaptive mutations that increase replication efficiency of HCV RNA is located in the central region of NS5A.[14–16] The membrane association of NS5A through its N-terminal transmembrane domain[55] and the interaction between NS5A and 5B[56] are essential for the RNA replication. Several cellular proteins interacting with NS5A have been identified, and human vesicle-associated membrane protein-associated protein A (h-VAP-A) is likely to play a key role in the RNA replication through the interacting with NS5A, as demonstrated by experiments using RNA interference and dominant-negative protein fragments.[49] In addition to its function related to the viral replication, NS5A appears to be implicated in resistance of HCV-infected cells to the antiviral effect of IFN. At least for some HCV isolates, NS5A interacts structurally and functionally with an IFN-induced protein kinase PKR, leading to the inhibition of the kinase function and hence blocking the translation reduction in the IFN-treated cells.[57,58]

NS5B is an RNA-dependent RNA polymerase (RdRp), a key enzyme involved in viral replication.[27,59–64] The RdRp activity of NS5B as well as its three-dimensional structure have been demonstrated using recombinant gene products prepared from a variety of expression systems. The optimal temperature, pH requirements and concentration of bivalent cations for the activity were found to be similar to those for the poliovirus 3D polymerase.[61] The X-ray studies have reported that while the enzyme has the typical right-handed "finger-palm-thumb" domains of the polymerase, extensive interactions of the fingers and thumb lead to a more fully enclosed active site tunnel, unlike other RdRp.[65–67] NS5B contains a hydrophobic domain at its C terminal 21 residues, which is a transmembrane segment. Recent studies have demonstrated that the C-terminal domain of NS5B appears to serve dual functions in the viral RNA replication, both through its role as a membrane anchor and through its involvement in RNA synthesis in a sequence-specific manner.[68,69]

RNA Replication Model

RNA replication of most RNA viruses involves certain intracellular membrane structures, including the ER,[70–72] Golgi,[73] endosomes and lysosomes.[74] Although NS5B protein has RdRp activity *in vitro*, its recombinant product alone is presumably short of the strict template specificity and fidelity, which are essential for the viral RNA synthesis. It is highly likely that other viral and/or host factors are important for conferring proper RNA replication and that the replication complexes (RCs) that are composed of NS5B and additional components required for modulating polymerase activity are involved in catalyzing HCV RNA synthesis during the replication process.

Several coprecipitation and immunostaining studies revealed that the newly synthesized HCV RNA was localized to distinct speckle-like structures, where all of the viral nonstructural proteins coexisted.[48] These distinct structures may be equivalent to the membranous webs, as reported by Gosert *et al*.[45] and described above. Expression of all structural and nonstructural proteins in the context of the entire HCV polyprotein induced similar membrane changes.[46] It is of interest that morphologically similar structures, termed sponge-like inclusions,[75] have been found by electron microscopy in liver cells of HCV-infected chimpanzees. Thus, the membranous web may comprise the HCV RC in infected cells.

Recently, several groups have succeeded in showing the *in vitro* replication activities of the HCV RCs in crude membrane fractions of cells harboring the subgenomic replicons.[76–80] These cell-free systems provide a valuable complement to the *in vitro* RdRp assays for biochemical dissection of HCV RNA replication and a useful source for isolation of viral RCs. The replication activity in the crude membrane fraction, which contains HCV ribonucleoprotein complexes associated with cellular membranes, is measured by incorporation of radiolabeled nucleotides into newly synthesized RNA *in vitro*, and the products can be resolved from replicative intermediates by native or denaturing gel electrophoresis. The RNA synthesis can be initiated in the absence of added negative-strand template RNA, suggesting that preinitiated template RNA copurifies with the RC.[77–79] Although the

newly synthesized single-strand RNA can be used as a template for a further round of double-strand RNA synthesis, no exogenous RNA can be served as a template for HCV RC preparation.[79] The added RNA templates might not access the active site of the HCV RCs due to sequestration by membranes. The HCV RCs contain both minus- and plus-strand RNAs,[77,78] with the plus-strand RNA being approximately 10-fold more abundant than the minus strand.[76] This is consistent with the replication process for other positive-strand RNA viruses, which produce a negative-sense replication intermediate in smaller amounts than that of the genomic RNA. It has also been reported that the cell-free replication activity increases at temperatures ranging from 25°C to 40°C, and divalent cations (Mn^{2+} and Mg^{2+}) can be used in the reaction.[78,79]

Shi *et al.*[48] have demonstrated that newly synthesized HCV RNA and the nonstructural proteins colocalized on distinct speckle-like structures in the cytoplasm of the replicon-containing cells. Membrane flotation analysis and replication assay have shown that the viral RNA and proteins were present in detergent-resistant membrane (DRM) structures, most likely a lipid-raft structure, and that RNA replication activity was detected even after treatment with detergent.[76,81] Lipid rafts are cholesterol- and sphingolipid-rich microdomains and are characterized by their detergent insolubility.[82–84] The structures are known to play a critical role in a number of biological processes such as regulators and organizing centers of signal transduction and membrane traffic pathways, including virus entry and assembly of, for example, influenza virus,[85–87] human immunodeficiency virus type-1 (HIV-1),[88–90] Ebola, Marburg virus,[91] enterovirus,[92] avian sarcoma and leukosis virus,[93] Coxsackie B virus, adenovirus,[94] measles virus,[1] and respiratory syncytial virus.[95] However, HCV may be the first example of the association of lipid rafts with viral RNA replication.

On the other hand, it has been widely thought that most of the HCV life cycle, including the protein processing and genome replication, takes place at the ER, where cholesterol-sphingolipid rafts are not assembled.[46,96–98] Several studies using the replicon system have indicated that the nonstructural proteins were associated with the ER.[97,99] Nevertheless, it is still possible that HCV nonstructural proteins that

are synthesized at the ER relocalize to lipid raft membranes when they are actively engaged in RNA replication. It has been shown by the membrane separation analysis that HCV nonstructural proteins existed both in the ER and the Golgi, but the activity of viral RNA replication was detected mainly in the Golgi fraction.[76,100] Further studies to elucidate where and how the HCV genome replicates in the infected cells are needed.

Following is a model for the HCV RC formation proposed by Aizaki *et al.*[76] HCV nonstructural proteins are processed from precursor polyprotein by the viral proteases and localized at the ER. A part of the nonstructural proteins and host factors interacted are transported to the Golgi, where they associate directly or indirectly, via protein-protein interactions, with lipid rafts (Fig. 1). NS4B may bind to lipid rafts first and then recruit other viral proteins consecutively to form RCs, since it appears that NS4B, but not other HCV proteins, associate with lipid rafts when expressed alone. A vesicle-associated protein hVAP-33, which possibly binds to NS5A and NS5B,[98] may also contribute to the formation of HCV RCs. Rafts harboring putative RCs may then be stabilized and combined to create DRM structures, presumably involved in the nature that NS4B can be oligomerized.[101] During the process of RC polymerization, HCV RNA is enclosed within the membrane complex. This process, which permits high local concentration of NS5B and the viral RNA, may be advantageous in order to confer template specificity and ensure RNA synthesis with necessary fidelity during the viral replication.

Host Factors Possibly Involved in HCV RNA Replication

Studies with the RNA replicons have demonstrated that the viral RNA level is highest in the growth phase of the cells and drops significantly when cells reach a confluent state, suggesting that HCV replication and/or translation are tightly linked to host cell metabolism.[99] Huh-7 cells in which adapted replicons are cured by treatment with IFN are able to yield cell populations that are more permissive for the replicon tested. Thus, it is likely that some interplay between

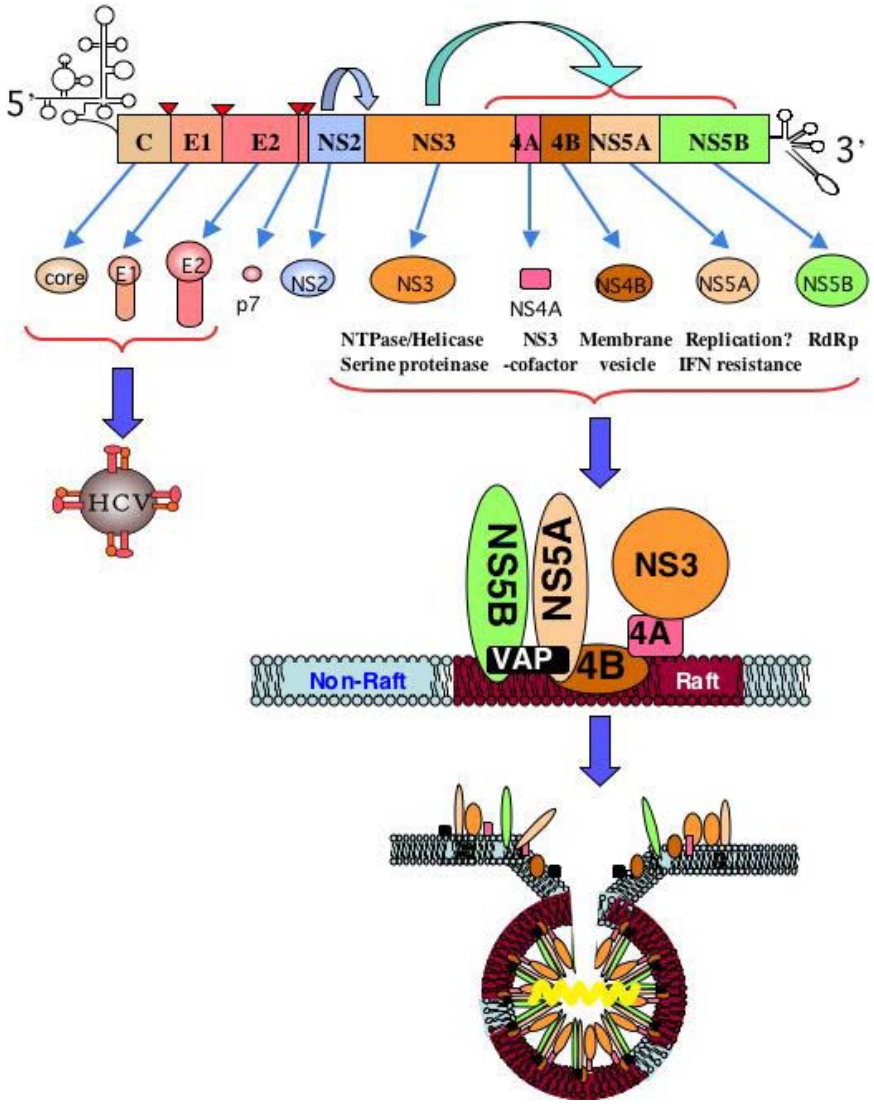


Fig. 1. Illustration of the HCV genome and replication complex. A schematic presentation of the HCV genome is given at the top. The individual cleavage products of the polyprotein are given in the second level down. The functions of the viral proteins are given in the third level down. A model of the formation of HCV replication complex on the lipid raft is shown at the bottom.

the cellular environment and particular adaptive mutations of the viral RNA is important for efficient RNA replication of HCV.

There is increasing experimental evidence that shows specific interactions between various cellular factors and HCV nonstructural proteins and/or RNA, possibly playing roles in HCV replication and/or translation. Several cellular proteins interacting with NS5A, such as hVAP-33,[98] growth factor receptor-bound protein 2 adaptor protein,[102] transcription factor SRCAP,[103] and karyopherin b3,[104] have been identified. Among them, it is of interest that hVAP-33, a SNARE-like protein, can interact with NS5B as well; NS5A binds to the C-terminus of hVAP-33, whereas NS5B binds to the N-terminus.[98] Down-regulation of hVAP-33, either through siRNA or expression of a truncated, dominant negative fragment of this protein, has lead to inhibition of HCV RNA replication, suggesting that protein-protein interactions among NS5A, NS5B and hVAP-33 are critical for formation of HCV RC and viral RNA replication (Fig. 1).[49] Another NS5B-interacting protein identified is a human RNA helicase, p68.[105] Expression of NS5B alone or of all the nonstructural proteins induced the redistribution of endogenous p68 from the nucleus to the cytoplasm. It has also been shown that knockdown of p68 reduces the negative-strand synthesis of the HCV genome, suggesting that the NS5B-p68 interaction and the subsequent relocalization of cellular proteins may serve to mediate HCV replication processes.

Cellular components binding to HCV RNA are also possibly involved in viral replication. Some candidates are La autoantigen, polyprimidine tract-binding protein (PTB), NS1-associated protein 1 (NSAP1), human heterogeneous nuclear ribonucleoprotein (hnRNP) L,[106] 40S ribosomal subunit protein,[107] and the gamma subunits of human eukaryotic initiation factors 2B (EIF2B-gamma) and 2 (EIF2-gamma).[108] La antigen, which is known to bind to several viral RNAs,[109–111] has been demonstrated to bind to the 5′-NTR of HCV RNA and stimulate HCV IRES-mediated translation.[112] PTB, a pre-mRNA splicing factor, is also known to bind to many viral RNAs at several different sites, such as the IRES of poliovirus[113] and hepatitis A virus.[114] It has been shown that PTB has an ability to

interact with the HCV 5′-UTR,[115] 3′-UTR[116–118] and core-protein-coding region of HCV RNA.[26] These reports have suggested that the interaction between the HCV RNA genome and PTB may provide both positive and negative regulation of the viral translation. A recent work has demonstrated that NSAP1, which is highly homologous to hnRNP R, enhances HCV IRES activity through an interaction with the core-protein-coding sequence of the viral genome.[119] Interestingly, the knockdown of endogenous La, PTB, or hVAP-33 with siRNA has been shown to efficiently inhibit HCV RNA replication in cultured cells.[120] It has been proposed that some translation regulatory factors may participate in viral RNA synthesis by either interacting directly with the viral RNA or associating with RdRp in other RNA viruses.[121] Thus, it may be possible that cellular proteins capable of binding to HCV RNA play roles in the viral transcription as well as the translation, although further meticulous experiments to separate viral transcription from viral translation are needed.

Future Perspectives

In addition to biochemical and structural analyses of HCV components carried out using heterologous expression systems, molecular biological studies with cell culture systems based on bicistronic, subgenomic and genomic HCV RNAs have provided us much information about the HCV lifecycle. Nevertheless, a number of questions concerning HCV replication still remain. For instance, cellular pathways involved in the formation of HCV RC are the focus of active investigation, but works to define host cellular factors using a variety of techniques are just beginning. While we are learning a lot from studying HCV replicon systems, current cell cultures harboring the selectable full-length HCV genome unfortunately still do not allow the production of viral particles and re-initiation of an infectious cycle. Development of a robust cell culture system that supports full replication of HCV would bring a further understanding of the viral life cycle.

References

1. Lohmann V, Korner F, Koch J, *et al.* (1999) Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285**: 110–113.
2. Andino R, Rieckhof GE, Achacoso PL, Baltimore D. (1993) Poliovirus RNA synthesis utilizes an RNP complex formed around the 5′-end of viral RNA. *EMBO J* **12**: 3587–3598.
3. Collis PS, O'Donnell BJ, Barton DJ, *et al.* (1992) Replication of poliovirus RNA and subgenomic RNA transcripts in transfected cells. *J Virol* **66**: 6480–6488.
4. Hagino-Yamagishi K, Nomoto A. (1989) *In vitro* construction of poliovirus defective interfering particles. *J Virol* **63**: 5386–5392.
5. Kaplan G, Racaniello VR. (1988) Construction and characterization of poliovirus subgenomic replicons. *J Virol* **62**: 1687–1696.
6. Liljestrom P, Garoff H. (1991) A new generation of animal cell expression vectors based on the Semliki Forest virus replicon. *Biotechnology (N Y)* **9**: 1356–1361.
7. Bredenbeek PJ, Frolov I, Rice CM, Schlesinger S. (1993) Sindbis virus expression vectors: packaging of RNA replicons by using defective helper RNAs. *J Virol* **67**: 6439–6446.
8. Johanning FW, Conry RM, LoBuglio AF, *et al.* (1995) A Sindbis virus mRNA polynucleotide vector achieves prolonged and high level heterologous gene expression *in vivo*. *Nucl Acids Res* **23**: 1495–1501.
9. Kamrud KI, Powers AM, Higgs S, *et al.* (1995) The expression of chloramphenicol acetyltransferase in mosquitoes and mosquito cells using a packaged Sindbis replicon system. *Exp Parasitol* **81**: 394–403.
10. Xiong C, Levis R, Shen P, *et al.* (1989) Sindbis virus: an efficient, broad host range vector for gene expression in animal cells. *Science* **243**: 1188–1191.
11. Khromykh AA, Westaway EG. (1997) Subgenomic replicons of the flavivirus Kunjin: construction and applications. *J Virol* **71**: 1497–1505.
12. McKnight KL, Lemon SM. (1996) Capsid coding sequence is required for efficient replication of human rhinovirus 14 RNA. *J Virol* **70**: 1941–1952.
13. Behrens SE, Grassmann CW, Thiel HJ, *et al.* (1998) Characterization of an autonomous subgenomic pestivirus RNA replicon. *J Virol* **72**: 2364–2372.

14. Blight KJ, Kolykhalov AA, Rice CM. (2000) Efficient initiation of HCV RNA replication in cell culture. *Science* **290**: 1972–1974.
15. Krieger N, Lohmann V, Bartenschlager R. (2001) Enhancement of hepatitis C virus RNA replication by cell culture-adaptive mutations. *J Virol* **75**: 4614–4624.
16. Lohmann V, Korner F, Dobierzewska A, Bartenschlager R. (2001) Mutations in hepatitis C virus RNAs conferring cell culture adaptation. *J Virol* **75**: 1437–1449.
17. Lohmann V, Hoffmann S, Herian U, *et al.* (2003) Viral and cellular determinants of hepatitis C virus RNA replication in cell culture. *J Virol* **77**: 3007–3019.
18. Yi M, Lemon SM. (2004) Adaptive mutations producing efficient replication of genotype 1a hepatitis C virus RNA in normal Huh7 cells. *J Virol* **78**: 7904–7915.
19. Ali S, Pellerin C, Lamarre D, Kukolj G. (2004) Hepatitis C virus subgenomic replicons in the human embryonic kidney 293 cell line. *J Virol* **78**: 491–501.
20. Kato T, Date T, Miyamoto M, *et al.* (2005) Nonhepatic cell lines HeLa and 293 support efficient replication of the hepatitis C virus genotype 2a subgenomic replicon. *J Virol* **79**: 592–596.
21. Kato N, Sugiyama K, Namba K, *et al.* (2003) Establishment of a hepatitis C virus subgenomic replicon derived from human hepatocytes infected *in vitro*. *Biochem Biophys Res Commun* **306**: 756–766.
22. Zhu Q, Guo JT, Seeger C. (2003) Replication of hepatitis C virus subgenomes in nonhepatic epithelial and mouse hepatoma cells. *J Virol* **77**: 9204–9210.
23. Pietschmann T, Lohmann V, Kaul A, *et al.* (2002) Persistent and transient replication of full-length hepatitis C virus genomes in cell culture. *J Virol* **76**: 4008–4021.
24. Rijnbrand R, Bredenbeek PJ, Haasnoot PC, *et al.* (2001) The influence of downstream protein-coding sequence on internal ribosome entry on hepatitis C virus and other flavivirus RNAs. *RNA* **7**: 585–597.
25. Friebe P, Lohmann V, Krieger N, Bartenschlager R. (2001) Sequences in the 5′ nontranslated region of hepatitis C virus required for RNA replication. *J Virol* **75**: 12047–12057.
26. Ito T, Lai MM. (1999) An internal polypyrimidine-tract-binding protein-binding site in the hepatitis C virus RNA attenuates translation, which is relieved by the 3′-untranslated sequence. *Virology* **254**: 288–296.

27. Kolykhalov AA, Agapov EV, Blight KJ, *et al.* (1997) Transmission of hepatitis C by intrahepatic inoculation with transcibed RNA. *Science* **277**: 570–574.
28. Tanaka T, Kato N, Cho M-J, Shimotohno K. (1995) A novel sequence found at the 3′ terminus of hepatitis C virus genome. *Biochem Biophys Res Commun* **215**: 744–749.
29. Yi M, Lemon SM. (2003) 3′-nontranslated RNA signals required for replication of hepatitis C virus RNA. *J Virol* **77**: 3557–3568.
30. Friebe P, Bartenschlager R. (2002) Genetic analysis of sequences in the 3′ nontranslated region of hepatitis C virus that are important for RNA replication. *J Virol* **76**: 5326–5338.
31. Jin L, Peterson DL. (1995) Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA helicase. *Arch Biochem Biophys* **323**: 47–53.
32. Kim DW, Gwack Y, Han JH, Choe J. (1995) C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity. *Biochem Biophys Res Commun* **215**: 160–166.
33. Suzich JA, Tamura JK, Palmer-Hill F, *et al.* (1993) Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. *J Virol* **67**: 6152–6158.
34. Tai CL, Chi WK, Chen DS, Hwang LH. (1996) The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). *J Virol* **70**: 8477–8484.
35. Tai CL, Pan WC, Liaw SH, *et al.* (2001) Structure-based mutational analysis of the hepatitis C virus NS3 helicase. *J Virol* **75**: 8289–8297.
36. Wolk B, Sansonno D, Krausslich HG, *et al.* (2000) Subcellular localization, stability, and trans-cleavage competence of the hepatitis C virus NS3-NS4A complex expressed in tetracycline-regulated cell lines. *J Virol* **74**: 2293–2304.
37. Barbato G, Cicero DO, Nardi MC, *et al.* (1999) The solution structure of the N-terminal proteinase domain of the hepatitis C virus (HCV) NS3 protein provides new insights into its activation and catalytic mechanism. *J Mol Biol* **289**: 371–384.
38. Tanji Y, Hijikata M, Satoh S, *et al.* (1995) Hepatitis C virus-encoded nonstructural protein NS4A has versatile functions in viral protein processing. *J Virol* **69**: 1575–1581.
39. Hugle T, Fehrmann F, Bieck E, *et al.* (2001) The hepatitis C virus nonstructural protein 4B is an integral endoplasmic reticulum membrane protein. *Virology* **284**: 70–81.

40. Kim JE, Song WK, Chung KM, *et al.* (1999) Subcellular localization of hepatitis C viral proteins in mammalian cells. *Arch Virol* **144**: 329–343.
41. Lundin M, Monne M, Widell A, *et al.* (2003) Topology of the membrane-associated hepatitis C virus protein NS4B. *J Virol* **77**: 5428–5438.
42. Selby MJ, Choo Q-L, Berger K, *et al.* (1993) Expression, identification and subcellular localization of the proteins encoded by the hepatitis C viral genome. *J Gen Virol* **74**: 1103–1113.
43. Florese RH, Nagano-Fujii M, Iwanaga Y, *et al.* (2002) Inhibition of protein synthesis by the nonstructural proteins NS4A and NS4B of hepatitis C virus. *Virus Res* **90**: 119–131.
44. Kato J, Kato N, Yoshida H, *et al.* (2002) Hepatitis C virus NS4A and NS4B proteins suppress translation *in vivo*. *J Med Virol* **66**: 187–199.
45. Piccininni S, Varaklioti A, Nardelli M, *et al.* (2002) Modulation of the hepatitis C virus RNA-dependent RNA polymerase activity by the non-structural (NS) 3 helicase and the NS4B membrane protein. *J Biol Chem* **277**: 45670–45679.
46. Egger D, Wolk B, Gosert R, *et al.* (2002) Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *J Virol* **76**: 5974–5984.
47. Gosert R, Egger D, Lohmann V, *et al.* (2003) Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *J Virol* **77**: 5487–5492.
48. Shi ST, Lee KJ, Aizaki H, *et al.* (2003) Hepatitis C virus RNA replication occurs on a detergent-resistant membrane that cofractionates with caveolin-2. *J Virol* **77**: 4160–4168.
49. Gao L, Aizaki H, He JW, Lai MM. (2004) Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft. *J Virol* **78**: 3480–3488.
50. Pawlotsky JM, Germanidis G. (1999) The non-structural 5A protein of hepatitis C virus. *J Viral Hepat* **6**: 343–356.
51. Coito C, Diamond DL, Neddermann P, *et al.* (2004) High-throughput screening of the yeast kinome: identification of human serine/threonine protein kinases that phosphorylate the hepatitis C virus NS5A protein. *J Virol* **78**: 3502–3513.
52. Ide Y, Tanimoto A, Sasaguri Y, Padmanabhan R. (1997) Hepatitis C virus NS5A protein is phosphorylated *in vitro* by a stably bound protein

kinase from HeLa cells and by cAMP-dependent protein kinase A-alpha catalytic subunit. *Gene* **201**: 151–158.
53. Kim J, Lee D, Choe J. (1999) Hepatitis C virus NS5A protein is phosphorylated by casein kinase II. *Biochem Biophys Res Commun* **257**: 777–781.
54. Reed KE, Gorbalenya AE, Rice CM. (1998) The NS5A/NS5 proteins of viruses from three genera of the family flaviviridae are phosphorylated by associated serine/threonine kinases. *J Virol* **72**: 6199–6206.
55. Brass V, Bieck E, Montserret R, *et al.* (2002) An amino-terminal amphipathic alpha-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. *J Biol Chem* **277**: 8130–8139.
56. Shimakami T, Hijikata M, Luo H, *et al.* (2004) Effect of interaction between hepatitis C virus NS5A and NS5B on hepatitis C virus RNA replication with the hepatitis C virus replicon. *J Virol* **78**: 2738–2748.
57. Gale MJ, Jr., Korth MJ, Katze MG. (1998) Repression of the PKR protein kinase by the hepatitis C virus NS5A protein: a potential mechanism of interferon resistance. *Clin Diagn Virol* **10**: 157–162.
58. Gale MJ, Jr., Korth MJ, Tang NM, *et al.* (1997) Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. *Virology* **230**: 217–227.
59. Al RH, Xie Y, Wang Y, Hagedorn CH. (1998) Expression of recombinant hepatitis C virus non-structural protein 5B in Escherichia coli. *Virus Res* **53**: 141–149.
60. Behrens S-E, Tomei L, DeFrancesco R. (1996) Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO J* **15**: 12–22.
61. Ishii K, Tanaka Y, Yap CC, *et al.* (1999) Expression of hepatitis C virus NS5B protein: characterization of its RNA polymerase activity and RNA binding. *Hepatology* **29**: 1227–1235.
62. Lohmann V, Korner F, Herian U, Bartenschlager R. (1997) Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity. *J Virol* **71**: 8416–8428.
63. Oh JW, Ito T, Lai MM. (1999) A recombinant hepatitis C virus RNA-dependent RNA polymerase capable of copying the full-length viral RNA. *J Virol* **73**: 7694–7702.
64. Yamashita T, Kaneko S, Shirota Y, *et al.* (1998) RNA-dependent RNA polymerase activity of the soluble recombinant hepatitis C virus

NS5B protein truncated at the C-terminal region. *J Biol Chem* **273**: 15479–15486.
65. Ago H, Adachi T, Yoshida A, *et al.* (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Structure Fold Des* **7**: 1417–1426.
66. Bressanelli S, Tomei L, Roussel A, *et al.* (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc Natl Acad Sci USA* **96**: 13034–13039.
67. Lesburg CA, Cable MB, Ferrari E, *et al.* (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat Struct Biol* **6**: 937–943.
68. Ivashkina N, Wolk B, Lohmann V, *et al.* (2002) The hepatitis C virus RNA-dependent RNA polymerase membrane insertion sequence is a transmembrane segment. *J Virol* **76**: 13088–13093.
69. Lee KJ, Choi J, Ou JH, Lai MM. (2004) The C-terminal transmembrane domain of hepatitis C virus (HCV) RNA polymerase is essential for HCV replication *in vivo*. *J Virol* **78**: 3797–3802.
70. Restrepo-Hartwig MA, Ahlquist P. (1996) Brome mosaic virus helicase- and polymerase-like proteins colocalize on the endoplasmic reticulum at sites of viral RNA synthesis. *J Virol* **70**: 8908–8916.
71. Schaad MC, Jensen PE, Carrington JC. (1997) Formation of plant RNA virus replication complexes on membranes: role of an endoplasmic reticulum-targeted viral protein. *EMB J* **16**: 4049–4059.
72. van der Meer Y, van Tol H, Locker JK, Snijder EJ. (1998) ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex. *J Virol* **72**: 6689–6698.
73. Shi ST, Schiller JJ, Kanjanahaluethai A, *et al.* (1999) Colocalization and membrane association of murine hepatitis virus gene 1 products and De novo-synthesized viral RNA in infected cells. *J Virol* **73**: 5957–5969.
74. Froshauer S, Kartenbeck J, Helenius A. (1988) Alphavirus RNA replicase is located on the cytoplasmic surface of endosomes and lysosomes. *J Cell Biol* **107**: 2075–2086.
75. Pfeifer U, Thomssen R, Legler K, *et al.* (1980) Experimental non-A, non-B hepatitis: four types of cytoplasmic alteration in hepatocytes of infected chimpanzees. *Virchows Arch B Cell Pathol Incl Mol Pathol* **33**: 233–243.
76. Aizaki H, Lee KJ, Sung VM, *et al.* (2004) Characterization of the hepatitis C virus RNA replication complex associated with lipid rafts. *Virology* **324**: 450–461.

77. Ali N, Tardif KD, Siddiqui A. (2002) Cell-free replication of the hepatitis C virus subgenomic replicon. *J Virol* **76**: 12001–12007.
78. Hardy RW, Marcotrigiano J, Blight KJ, *et al.* (2003) Hepatitis C virus RNA synthesis in a cell-free system isolated from replicon-containing hepatoma cells. *J Virol* **77**: 2029–2037.
79. Lai VC, Dempsey S, Lau JY, *et al.* (2003) *In vitro* RNA replication directed by replicase complexes isolated from the subgenomic replicon cells of hepatitis C virus. *J Virol* **77**: 2295–2300.
80. Miyanari Y, Hijikata M, Yamaji M, *et al.* (2003) Hepatitis C virus nonstructural proteins in the probable membranous compartment function in viral genome replication. *J Biol Chem* **278**: 50301–50308.
81. Waris G, Sarker S, Siddiqui A. (2004) Two-step affinity purification of the hepatitis C virus ribonucleoprotein complex. *RNA* **10**: 321–329.
82. Simons K, Ikonen E. (1997) Functional rafts in cell membranes. *Nature* **387**: 569–572.
83. Simons K, Ikonen E. (2000) How cells handle cholesterol. *Science* **290**: 1721–1726.
84. Simons K, Toomre D. (2000) Lipid rafts and signal transduction. *Nat Rev Mol Cell Biol* **1**: 31–39.
85. Barman S, Ali A, Hui EK, *et al.* (2001) Transport of viral proteins to the apical membranes and interaction of matrix protein with glycoproteins in the assembly of influenza viruses. *Virus Res* **77**: 61–69.
86. Scheiffele P, Rietveld A, Wilk T, Simons K. (1999) Influenza viruses select ordered lipid domains during budding from the plasma membrane. *J Biol Chem* **274**: 2038–2044.
87. Zhang J, Pekosz A, Lamb RA. (2000) Influenza virus assembly and lipid raft microdomains: a role for the cytoplasmic tails of the spike glycoproteins. *J Virol* **74**: 4634–4644.
88. Ding L, Derdowski A, Wang JJ, Spearman P. (2003) Independent segregation of human immunodeficiency virus type 1 Gag protein complexes and lipid rafts. *J Virol* **77**: 1916–1926.
89. Manes S, del Real G, Lacalle RA, *et al.* (2000) Membrane raft microdomains mediate lateral assemblies required for HIV-1 infection. *EMBO Rep* **1**: 190–196.
90. Ono A, Freed EO. (2001) Plasma membrane rafts play a critical role in HIV-1 assembly and release. *Proc Natl Acad Sci USA* **98**: 13925–13930.
91. Bavari S, Bosio CM, Wiegand E, *et al.* (2002) Lipid raft microdomains: a gateway for compartmentalized trafficking of Ebola and Marburg viruses. *J Exp Med* **195**: 593–602.

92. Stuart AD, Eustace HE, McKee TA, Brown TD. (2002) A novel cell entry pathway for a DAF-using human enterovirus is dependent on lipid rafts. *J Virol* **76**: 9307–9322.
93. Narayan S, Barnard RJ, Young JA. (2003) Two retroviral entry pathways distinguished by lipid raft association of the viral receptor and differences in viral infectivity. *J Virol* **77**: 1977–1983.
94. Ashbourne Excoffon KJ, Moninger T, Zabner J. (2003) The coxsackie B virus and adenovirus receptor resides in a distinct membrane microdomain. *J Virol* **77**: 2559–2567.
95. Brown DA, Rose JK. (1992) Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. *Cell* **68**: 533–544.
96. Hijikata M, Mizushima H, Tanji Y, *et al.* (1993) Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus. *Proc Natl Acad Sci USA* **90**: 10773–10777.
97. Mottola G, Cardinali G, Ceccacci A, *et al.* (2002) Hepatitis C virus nonstructural proteins are localized in a modified endoplasmic reticulum of cells expressing viral subgenomic replicons. *Virology* **293**: 31–43.
98. Tu H, Gao L, Shi ST, *et al.* (1999) Hepatitis C virus RNA polymerase and NS5A complex with a SNARE-like protein. *Virology* **263**: 30–41.
99. Pietschmann T, Lohmann V, Rutter G, *et al.* (2001) Characterization of cell lines carrying self-replicating hepatitis C virus RNAs. *J Virol* **75**: 1252–1264.
100. Choi J, Lee KJ, Zheng Y, *et al.* (2004) Reactive oxygen species suppress hepatitis C virus RNA replication in human hepatoma cells. *Hepatology* **39**: 81–89.
101. Dimitrova M, Imbert I, Kieny MP, Schuster C. (2003) Protein-protein interactions between hepatitis C virus nonstructural proteins. *J Virol* **77**: 5401–5414.
102. Tan SL, Nakao H, He Y, *et al.* (1999) NS5A, a nonstructural protein of hepatitis C virus, binds growth factor receptor-bound protein 2 adaptor protein in a Src homology 3 domain/ligand-dependent manner and perturbs mitogenic signaling. *Proc Natl Acad Sci USA* **96**: 5533–5538.
103. Ghosh AK, Majumder M, Steele R, *et al.* (2000) Hepatitis C virus NS5A protein modulates transcription through a novel cellular transcription factor SRCAP. *J Biol Chem* **275**: 7184–7188.
104. Chung KM, Lee J, Kim JE, *et al.* (2000) Nonstructural protein 5A of hepatitis C virus inhibits the function of karyopherin beta3. *J Virol* **74**: 5233–5241.

105. Goh PY, Tan YJ, Lim SP, *et al.* (2004) Cellular RNA helicase p68 relocalization and interaction with the hepatitis C virus (HCV) NS5B protein and the potential role of p68 in HCV RNA replication. *J Virol* **78**: 5288–5298.
106. Hahm B, Kim YK, Kim JH, *et al.* (1998) Heterogeneous nuclear ribonucleoprotein L interacts with the 3′-border of the internal ribosomal entry site of hepatitis C virus. *J Virol* **72**: 8782–8788.
107. Fukushi S, Okada M, Kageyama T, *et al.* (1999) Specific interaction of a 25-kilodalton cellular protein, a 40S ribosomal subunit protein, with the internal ribosome entry site of hepatitis C virus genome. *Virus Genes* **19**: 153–161.
108. Kruger M, Beger C, Li QX, *et al.* (2000) Identification of eIF2Bgamma and eIF2gamma as cofactors of hepatitis C virus internal ribosome entry site-mediated translation using a functional genomics approach. *Proc Natl Acad Sci USA* **97**: 8566–8571.
109. Kurilla MG, Cabradilla CD, Holloway BP, Keene JD. (1984) Nucleotide sequence and host La protein interactions of rabies virus leader RNA. *J Virol* **50**: 773–778.
110. Kurilla MG, Keene JD. (1983) The leader RNA of vesicular stomatitis virus is bound by a cellular protein reactive with anti-La lupus antibodies. *Cell* **34**: 837–845.
111. Wilusz J, Kurilla MG, Keene JD. (1983) A host protein (La) binds to a unique species of minus-sense leader RNA during replication of vesicular stomatitis virus. *Proc Natl Acad Sci USA* **80**: 5827–5831.
112. Ali N, Pruijn GJ, Kenan DJ, *et al.* (2000) Human La antigen is required for the hepatitis C virus internal ribosome entry site-mediated translation. *J Biol Chem* **275**: 27531–27540.
113. Hellen CU, Pestova TV, Wimmer E. (1994) Effect of mutations downstream of the internal ribosome entry site on initiation of poliovirus protein synthesis. *J Virol* **68**: 6312–6322.
114. Venkatramana M, Ray PS, Chadda A, Das S. (2003) A 25 kDa cleavage product of polypyrimidine tract binding protein (PTB) present in mouse tissues prevents PTB binding to the 5′-untranslated region and inhibits translation of hepatitis A virus RNA. *Virus Res* **98**: 141–149.
115. Ali N, Siddiqui A. (1995) Interaction of polypyrimidine tract-binding protein with the 5′ noncoding region of the hepatitis C virus RNA genome and its functional requirement in internal initiation of translation. *J Virol* **69**: 6367–6375.

116. Ito T, Lai MM. (1997) Determination of the secondary structure of and cellular protein binding to the 3′-untranslated region of the hepatitis C virus RNA genome. *J Virol* **71**: 8698–8706.
117. Murakami K, Abe M, Kageyama T, *et al.* (2001) Down-regulation of translation driven by hepatitis C virus internal ribosomal entry site by the 3′ untranslated region of RNA. *Arch Virol* **146**: 729–741.
118. Tsuchihara K, Tanaka T, Hijikata M, *et al.* (1997) Specific interaction of polypyrimidine tract-binding protein with the extreme 3′-terminal structure of the hepatitis C virus genome, the 3′X. *J Virol* **71**: 6720–6726.
119. Kim JH, Paek KY, Ha SH, *et al.* (2004) A cellular RNA-binding protein enhances internal ribosomal entry site-dependent translation through an interaction downstream of the hepatitis C virus polyprotein initiation codon. *Mol Cell Biol* **24**: 7878–7890.
120. Zhang J, Yamada O, Sakamoto T, *et al.* (2004) Down-regulation of viral replication by adenoviral-mediated expression of siRNA against cellular cofactors for hepatitis C virus. *Virology* **320**: 135–143.
121. Lai MM. (1998) Cellular factors in the transcription and replication of viral RNA genomes: a parallel to DNA-dependent RNA transcription. *Virology* **244**: 1–12.
122. Manie SN, Debreyne S, Vincent S, Gerlier D. (2000) Measles virus structural components are enriched into lipid raft microdomains: a potential cellular location for virus assembly. *J Virol* **74**: 305–311.

Chapter 8

Structure and Dynamics in Viral RNA Packaging

Thorsten Dieckmann[*,†] *and Marta Zumwalt*[†]

Interactions between RNA and proteins play a central role in numerous crucial cellular processes ranging from protein synthesis to RNA editing and translational control. In the course of the viral lifecycle molecular-level interactions between the viral protein and RNA components are vital for the control of the infection and correct assembly of newly produced virus particles. Many of these interactions are today the focus of research in order to develop new strategies for anti-viral treatments. One area of interest is the recognition process between viral coat or capsid proteins and the genomic RNA, which is crucial in order to ensure the packaging of the required genomic material. Here we present an overview of the molecular-level interactions that control this process in three different types of RNA viruses.

RNA-Protein Interactions in Viral RNA Packaging

The lifecycle of RNA viruses depends in many steps on molecular-level interactions between the viral genomic RNA and various cellular and viral proteins. These include the translation or reverse transcription of the RNA by enzymes, the interactions of regulatory elements, and the packaging of the correct set of viral genomic RNA into the

*Corresponding author. Email: tdieckma@sciborguwaterloo.ca.
†Deparment of Chemistry, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada.

newly produced virus particles. Recent progress in X-ray crystallography, NMR spectroscopy and electron microscopy has provided a wealth of structural data about how RNA and viral coat or capsid proteins recognize each other with high specificity and affinity to result in a highly accurate and efficient packaging process despite the "chemical chaos" found inside of the host cell. Here we present an overview of the atomic-level interactions between protein and RNA that drive this process in three different systems: the MS2 bacteriophage, the alfalfa mosaic virus, and retroviruses.

Packaging of Viral RNA

The process of incorporating the newly produced viral genome into the assembling particle is called packaging. This crucial step during the viral lifecycle follows one of two general mechanisms termed concerted and sequential assembly. During concerted assembly the building blocks of the viral protein shell can only assemble in a productive manner in the presence of the genomic nucleic acid. On the other hand, the genomic nucleic acid is inserted into a preformed protein shell in the course of the sequential assembly pathway.

During the packaging process, viral genomic RNA must be distinguished from cellular nucleic acids present in the same cellular compartment in which the assembly occurs. This includes cellular messenger RNAs, transfer RNA, ribosomal RNAs, and the degradation products thereof. In order to achieve this, the building blocks of the viral protein shell must be able to recognize their correct partner(s) in amid many similar nucleic acid molecules. For example, in the case of many retroviruses, the genomic RNA exists only in amounts of less than 1% of the total cellular RNA found in the cytoplasm. Nevertheless, the vast majority of newly produced virus particles contain the correct RNA. This remarkable degree of specificity is accomplished by molecular-level interactions between distinct sequences and structures within the coat or capsid proteins and the genomic RNA. The unique sequences or structures found in the viral genome are termed packaging signals. The size, secondary structures, and sequences found in RNA viruses range from simple hairpin

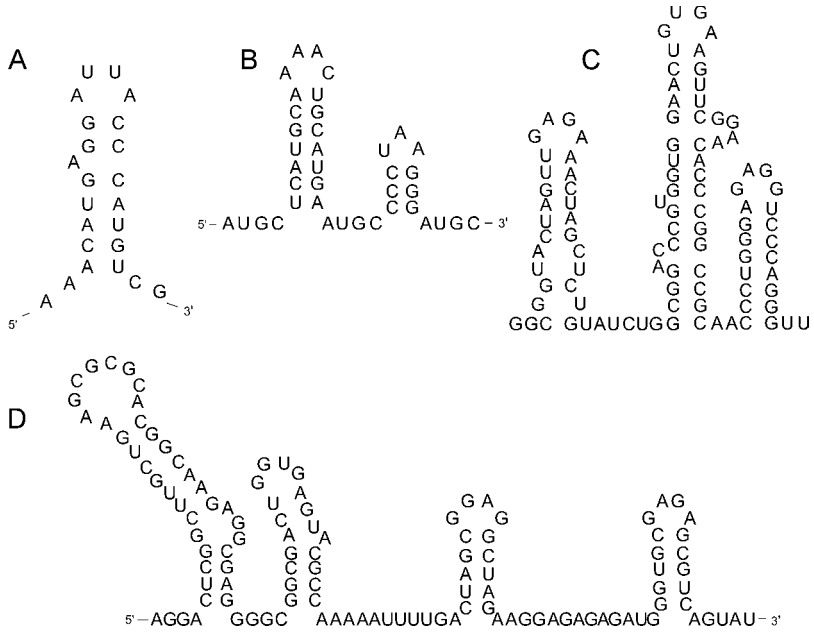


Fig. 1. Packaging signal sequences of MS2 phage (A), AMV (B), MMLV (C) and HIV-1 (D).

structures of 20–30 nucleotides to hundreds of nucleotides distributed over several sites within the genome (Fig. 1). In addition to this direct recognition of RNA and protein viruses that contain a multiple-part genome also employ RNA-RNA interactions in order to facilitate the accurate assembly of the new particles. An example of this can be found in retro-viruses such as human immunodeficiency virus (HIV), which package their genome as a dimer.

Specific packaging signals are a requirement for the packaging of a particular genome; however, their simple existence does not always guarantee packaging. One limiting factor is the total size of the RNA containing the packaging signal when compared to the space available inside the mature particle. The highest accuracy and efficiency of packaging can only be achieved with a genomic RNA that "fits" into the protein shell. This sets a natural size limit for the use of viral vectors as a means of introducing foreign RNA into cells. Another factor

that greatly affects packaging accuracy is the coupling of encapsidation and synthesis of the genome as found in many (−) strand RNA viruses, for example the influenza A virus. In a similar manner, sequestration of the viral RNA and capsid/coat protein in vesicles as found in cells infected with polio virus greatly improve packaging efficiency by increasing the local concentration of the correct components while keeping competing cellular material at low concentrations.

The following sections will provide an overview of the molecular-level interactions that control packaging in four well-studied model systems: The MS2 bacteriophage, the alfalfa mosaic virus (AMV), and the retroviruses HIV and moloney murine leukemia virus (MMV).

The MS2 Phage

The MS2 phage is a group I bacteriophage with a positive single-stranded RNA genome comprising 3569 nucleotides. The major gene products that are expressed upon viral infection of *E. coli* cells are the 393-amino acid maturation protein, a 60.7 kDa replicase subunit, the 13.75 kDa phage coat protein and the 75-amino acid lysis protein.

During the packaging of the genomic RNA a signal sequence within the genomic RNA forms a specific complex with a coat protein dimer as the intra-cellular concentration of coat protein increases. The RNA-protein complex acts as a repressor of the translation of the viral replicase gene by binding to an element in the initiation codon. This is a key interaction in the regulation of virus assembly. The MS2 system has been widely used to study sequence-specific interactions between RNA and protein using biochemical and biophysical methods.[1–5]

The RNA Packaging Signal

The coat protein of the MS2 phage interacts specifically with a 19-nucleotide RNA hairpin that is located at the 5′ end of the viral replicase gene. The hairpin consists of a seven-base pair stem closed by a four-nucleotide loop. The stem is interrupted by an unpaired nucleotide between base pairs five and six (Fig. 1). The stem sequence includes the Shine-Dalgarno sequence and initiation codon of the

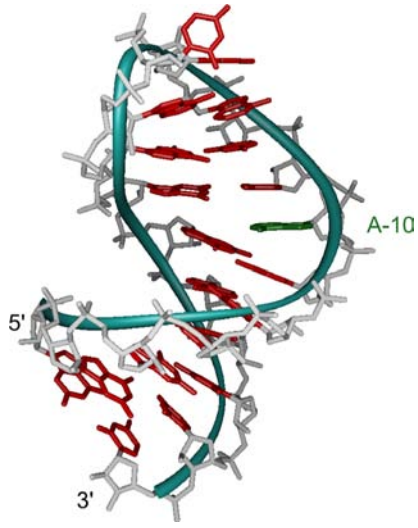


Fig. 2. Three-dimensional structure of the MS2 packaging signal RNA (1DOU).

viral replicase gene; however, biochemical studies showed that neither of these is involved in direct interactions with the coat protein.[6] The structure of the isolated RNA stem-loop has been determined by NMR spectroscopy for a closely related virus, the bacteriophage R17.[7] The structure shows that the unpaired nucleotide (A-10, Fig. 2) is intercalated in the stem. The entire segment of the sequence from G-11 to C-3 shows signs of a high degree of flexibility. The position of the bases in the loop is not very well defined. This type of partially unstructured or highly dynamic RNA binding site is a feature found in several RNA protein complexes.[8–11]

The Coat Protein

The 129 amino acids comprising MS2 coat protein form dimers at concentrations below 1.0 μM. At higher concentrations these dimers associate to form empty protein shells that are very similar in structure to native virus particles.[5,12] Each of these particles contains 90 coat protein dimers. Each monomer of the coat protein contains seven β-strands and two α-helices. Five of the β-strands are arranged

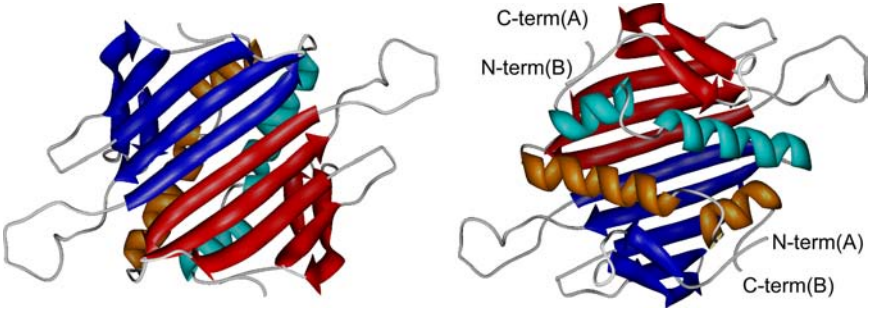


Fig. 3. Structure of the MS2 coat protein dimer. View from inside (left) and outside (right) the virus particle. One monomer is shown in red and orange, the other one in blue and cyan (1MSC).

in a β-meander topology (Fig. 3). The fundamental building block of the viral protein shell is the compact protein dimer that exists in two distinct but similar conformations. In the dimer 10 of the β-strands form a twisted β-sheet that points toward the interior of the virus and thus the RNA, whereas the α-helices are on the surface of the particle (Fig. 3).[5,12] The surface of the β-sheet has an excess of positively charged amino acids, resulting in an overall positive electrostatic environment ideal for interactions with the poly-anionic RNA.

The RNA-Coat Protein Complex

The structure of the complex between the MS2 coat protein and the packaging signal RNA hairpin has been extensively studied by X-crystallography[4,5,12–16] and biochemical approaches.[2,17–20] In the virus the 10-stranded β-sheet of the protein dimer faces toward the RNA-filled interior of the particle. The RNA hairpin binds to this sheet and contacts both subunits of the dimer (Fig. 4). The helical axis of the short stem is oriented almost parallel to the strands of the sheet. The 5′ side of the stem and part of the loop are facing the protein. The total interaction surface between RNA and protein is approximately 1700 Å^2. Twenty-one amino-acid residues in the dimer are directly involved in interactions with the RNA (Fig. 4). Of these, 13 contact the bases of nucleotides −4, −5, and −10, which have been shown to

Fig. 4. Structure of the MS2 coat protein — RNA complex.

be crucial for specific binding. Additional backbone contacts are largely with phosphates. However, there is one hydrogen bond between the 2′OH of U-5 and the carboxylate group of Glu 63.[17] The phosphate interactions are mainly mediated by hydrogen bonds and ion pairs with lysine and arginine side chains.

Among the base-specific contacts between RNA and protein, the interactions with the bulged A-10 and A-4 in the loop are especially intriguing. Two very similar binding pockets, one on each subunit of the protein dimer, are used to recognize these two bases and are partly responsible for the high specificity of the interaction. In both pockets the base is sandwiched between the aliphatic part of the side chain of lysine 61 and the methyl group of threonine 59 on one side, and the side chain of valine 29 on the other (Fig. 4). Even though the three-dimensional structure of the binding pockets is quite similar, the two adenine bases are bound differently. A-4 is oriented such that its N1 is bound to Serine 47, while N6 and N7 are bound to threonine 45. On the other hand, A-10 is bound to Ser 47 via its N3, while its N1 and N6 form hydrogen bonds with threonine 45. These differences are caused by the asymmetry. The binding pockets are highly conserved in related phages.[21,22]

The uridine in the loop at position −5 is stacking with the side chain of tyrosine 85, a type of interaction frequently seen in RNA protein interactions.[23] The hydroxyl group of the tyrosine is involved in a hydrogen bond to the U-5 phosphate group. The specificity for a pyrimidine in the position seems to be established by a hydrogen bond between asparagines 87 and the O2 of the base (Fig. 4). The adenine in position −7 is stacked below the uridine base in position −5.

A comparison between the RNA structure in its free and bound form illustrates what has become a common theme in RNA-protein interactions: Whereas the protein adjusts its structure only slightly in the process of complex formation, the RNA undergoes a significant structural transition often including the folding of previously unstructured elements. In the case of the MS2 packing signal hairpin, the adenine in position −10 — which is stacked inside the stem in the free RNA — is bulged out of the helix and becomes one of the key players in conferring specificity. Furthermore, there is no evidence for the stacking of the loop nucleotide at position −5 on top of the adenine at position −7 as seen in the complex. The whole loop segment of the free RNA is not well-structured and shows signs of internal dynamics.

Alfalfa Mosaic Virus

The alfalfa mosaic virus (AMV) is a positive-sense RNA plant virus. Its genome consists of three genomic RNAs (RNAs 1, 2, and 3) and one subgenomic RNA (RNA 4). Each of the RNAs are separately recognized and packaged.[24] RNAs 1 and 2 encode proteins with replicase functions,[25] the viral movement protein (required for cell-to-cell virus movement) is encoded by RNA 3, and the subgenomic RNA 4 (transcribed from the negative-sense of RNA 3) encodes the AMV coat protein.[26]

The encapsidation of the genomic RNAs is controlled by specific, high-affinity signal sequences that comprise 39 nucleotides and are located in the 3′-untranslated regions of all four RNAs (Fig. 1).[27–29] This minimal sequence is competent for binding either full-length AMV coat protein or 26-amino acid peptides that contain the N-terminal RNA binding domain of the coat protein.[27] The Pro-Thr-x-Arg-Ser-x-x-Thy

sequence of this RNA binding domain is conserved among AMV and related viruses (ilarviruses).[27] It has been demonstrated that the arginine at the center of this sequence (position 17 in the AMV coat protein) is crucial for RNA binding.[27–31] Circular dichroism studies and the observation that crystallization of the coat protein requires proteolytic cleavage of the N-terminus suggest that the N-terminal RNA binding domain of the AMV coat protein is unstructured in the absence of RNA.[32–34]

The recent co-crystal structure of the RNA signal sequence and a 26-amino acid peptide based on the AMV coat protein N-terminus provides a first glance at how RNA and protein interact in this virus to ensure correct packaging.[35] The structure also reveals the significance of the AUGC repeats found in the 3′ UTRs of all AMV genomic RNAs.

The structure illustrates that the RNA and the coat protein N-terminus co-fold to form a novel structure in which the peptide and the RNA adopt conformations that substantially differ from their unbound forms (Fig. 5). The core of the peptide that contains the recognition sequence PTxRSxxY takes on an α-helical structure followed by an L-turn, which puts all crucial amino acids in direct contact with the RNA. The 29-nucleotide RNA segment undergoes a conformational transition that leads to the formation of base pairs between neighboring AUGC repeats (Fig. 5). In this unusual arrangement, the uridine in position 844 and the cytosine in position 846 pair with adenine 865 and guanine 867, respectively. The resulting G:C pair is a Watson–Crick type base pair; however, the A:U pair adopts the non-canonical reverse Watson–Crick topology (Fig. 5). The two new base pairs form a continuation of the first stem (nt 847–864) and an identical arrangement is seen at the base of stem 2. The formation of these base pairs significantly distorts the structure of the AMV RNA. The AUGC repeat that is located in between the two stem-loops (nt 865–868) participates in base pairing interactions at the bottom of both stems. A direct consequence of this is that alternating bases in the AUGC sequence are pointing into opposite directions. The resulting stacking arrangement of the bases introduces a kink into the RNA backbone.

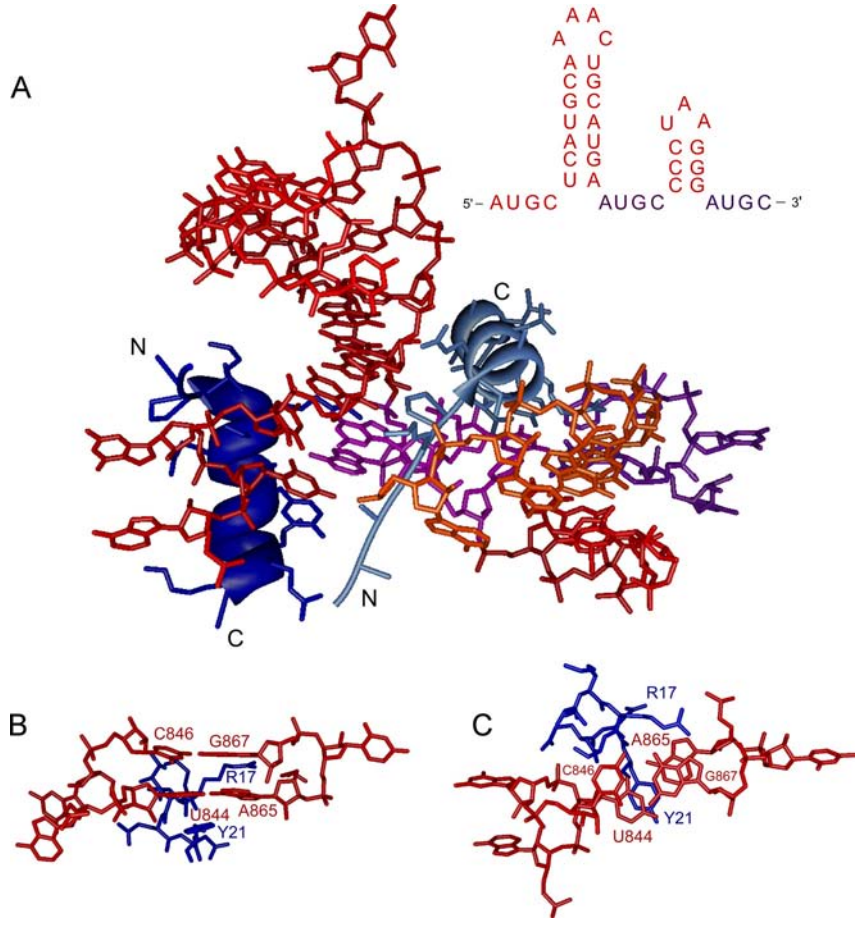


Fig. 5. Overview of the structure of the AMV coat protein-RNA complex (A) and detailed interactions of the interhelical (B) and 3´-terminal AUGC repeats (C) (1XOK).

The amino acids found in the conserved RNA binding motif of AMV coat protein stabilize the AUGC base pairing interactions via hydrogen bonds and stacking interactions. The crucial arginine residue (Arg 17) is located at a central position within the structure and inserts into the base stack at the major groove. Its position allows it to form hydrogen bonds with the bases of G867 (G17 of peptide 1) and G880 (G17 of peptide 2). Tyrosine 21 stacks below the A:U pair at the base

of each stem and thus effectively caps the structure (Fig. 5). Proline 14 introduces a sharp turn into the peptide chain that forms an L-shape and allows the N-terminal part of peptide 2 to follow the shape of the RNA surface, and in the case of peptide 1 to provide additional contacts with the RNA backbone. This conformation is stabilized by hydrogen bonds from the side chain hydroxyl groups and main chain amides of threonine 15 and serine 18.

The mode of recognition between RNA and AMV coat protein again demonstrates the importance of a dynamic RNA that facilitates folding in the presence of the protein partner. However, in this structure the dynamics go a step further and also include the protein because both the RNA signal sequence and the RNA binding domain of the coat protein undergo significant structural rearrangement during co-folding. This system also illustrates that the proposed secondary structure of the RNA signal sequence can be misleading. Even though the sequence folds into two stem-loop structures, similar to the MS2 signal sequence, the interaction with the AMV coat protein ignores the loops and takes place at the bottom of the stem with the AUGC repeats playing a key role.

Retroviruses

All retroviruses encode a gag polyprotein, which is produced in the host cell during the late stages of the infection cycle. This protein directs the encapsidation of two copies of the unspliced viral genome during viral assembly and budding. Gag is cleaved into the matrix protein (MA), the capsid protein (CA), and the nucleocapsid protein (NC) by the viral protease during viral maturation. The nucleocapsid protein associates with the viral RNA molecules that are encapsidated in the viral core particle.[36]

The Nucleocapsid Protein

The genome recognition in most retroviruses seems to be primarily mediated by the NC domain of the gag precursor polyprotein. Structures of several NC proteins have been determined in the past

decade, providing a wealth of detail about the protein side of this crucial interaction. The efforts are largely driven by the global endeavor to combat acquired immunodeficiency syndrome (AIDS), whose causative agent is the HIV retrovirus.[37–42] However, related retroviruses like the moloney murine leukemia virus (MMLV),[43] mouse mammary tumor virus (MMTV)[44] and Mason-Pfitzer monkey virus (MPMV)[45] were also extensively studied. All of these proteins have in common that they contain one or two copies of a CCHC-type zinc knuckle motif with the conserved sequence Cys-X_2-Cys-X_4-His-X_4-Cys, where X stands for a conservatively substituted amino acid. This sequence/structure motif is directly involved in RNA binding, as demonstrated by the fact that mutations that interfere with Zn binding greatly inhibit or abolish genome packaging.[46–50] All of these zinc knuckles contain a regular pattern of conservatively substituted hydrophobic and aromatic residues that form a cleft on the surface of the structure. Substitutions of these amino acids result in altered RNA packaging specificity, indicating that they are involved in sequence-specific RNA-protein interactions.[51]

The NC protein of human immunodeficiency virus type-1 (HIV-1) contains two such zinc knuckle domains, which are separated by a short linker sequence (RAPRKKG). The amino-terminal domain is essential for genome recognition.[52] The carboxy-terminal domain plays additional roles in viral assembly and during the early stages of viral infection.[53,54] The two domains adopt similar three-dimensional structures with the amino-terminal X-C-X-X-C-X-G-X sequence folding into a metal-coordinating reverse turn that has been termed "rubredoxin knuckle" because of its similarity with the structure found in the iron-coordinating domain of rubredoxin.[42] The subsequent residues form a loop followed by a 3_{10}-helix (Fig. 6).[41,42] The linker sequence shows little evidence of structure; however, NMR spectroscopic studies revealed transient interactions between the two zinc-knuckle domains.[37,38] The overall structure of the NC protein can be best described by a rapid equilibrium between structures in which the knuckles interact with each other and where they are independent from each other. In light of other RNA-protein interactions, it is likely

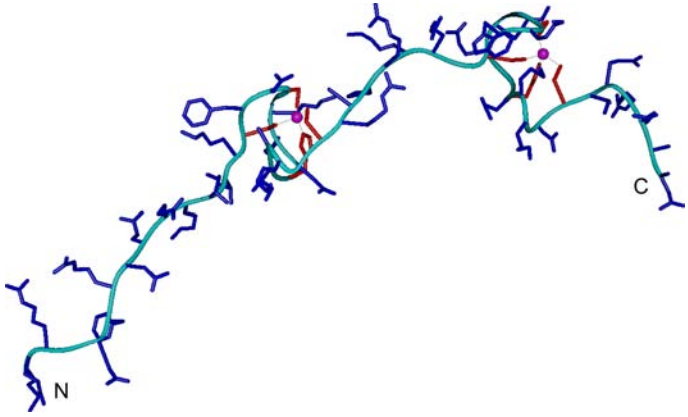


Fig. 6. Solution structure of the HIV-1 nucleocapsid domain zinc knuckles (1AAF).

that this internal flexibility is important for binding of the NC protein to its nucleic acid targets, also known as adaptive binding.

The HIV-1 RNA Packaging Signal

Genome recognition and packaging in retroviruses is mediated by interactions between the NC protein and a long stretch of nucleotides in the unspliced viral genome (about 120 nucleotides in the case of HIV-1). This sequence has become known as the Ψ-site and is generally located between the 5´-long terminal repeat and the gag initiation codon.[46,55,56] In all known cases the Ψ-sites overlap with the RNA dimer initiation elements that ensure packaging of two copies of the unspliced genomic RNA (see below).[57] The secondary structures of the Ψ-site RNAs of HIV-1 and MMLV (Fig. 1) have been studied extensively by molecular biology and computational approaches.[58–65] The motifs contain several stem-loop structures, four in the case of HIV-1 (termed SL1 to SL4) and three for MMLV, connected by short, single-stranded linkers of 4–15 nucleotides. Several high-resolution structures have been obtained for RNA constructs based on these secondary structures.[66]

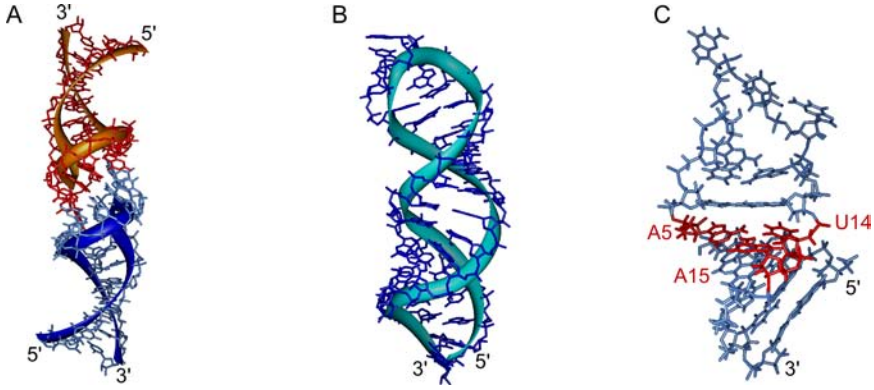


Fig. 7. Structure of the HIV-1 Ψ-site RNAs. Shown are the kissing hairpin structure formed by two SL1 RNAs (A), the monomeric form of SL1 (B) and the structure of SL2 with its AUA triplet (C) (1N8X, 1ESY).

The stem-loop I (SL1) of the HIV-1 Ψ-site contains the dimer initiation element, and it has been found that this portion of the RNA can self-associate to form either "kissing hairpin" structures or duplexes depending on the solution conditions.[67,68] The duplex structure consists of an extended A-form helix with two bulged adenines (Fig. 7) and indicates the capability of SL1 to adopt an alternate duplex structure rather that the expected stem-loop structure. This structural feature is likely to be crucial for packaging of the dimeric genome. The structure of the upper part of SL1 has been shown to form kissing hairpin dimers via base pairing of the six base palindromes GCGCGC and a distinctive pattern of inter-strand base stacking interactions of nucleotides in the stem-loop junctions (Fig. 7). A solution structure of the complete stem and bulge segment of SL1 was obtained by substituting the wild-type loop with an extra stable tetraloop (GAGA) and thus preventing dimer formation.[69] The resulting structure shows two helical regions separated by a bulge that results in an angle of approximately 25° between the upper and lower stem. The bulge consists of a G:A mismatch that stacks with the upper stem and two guanines that stack between the mismatch and the lower stem (Fig. 7).

The structure of HIV-1 SL2, which binds specifically to the NC protein, was determined by NMR spectroscopy.[70] The molecule adopts a stem-loop structure in which the loop region is largely disordered. The stem adopts a base pairing pattern that differs from the predicted secondary structure. Instead of the predicted A:U base and a bulged out A nucleotide, the structure contains an A-U-A base triple platform (Fig. 7).

SL3 also adopts the expected stem-loop arrangement. The stem is a Watson–Crick paired A-form helix and the loop is flexible, yet shows some preferred stacking patterns for the bases at the loop-stem interface.

The structure of a complete Ψ-site RNA was recently determined for MMLV.[71] In this 101-nucleotide construct the three individual stem-loops (SL-B, SL-C, and SL-D) of the core encapsidation signal are well defined. SL-C and SL-D are stacking end-to-end on each other and SL-B is connected to this longer segment via a conserved flexible linker (Fig. 8).

Complexes between the HIV-1 Ψ-site RNAs and the NC Protein

The four stem-loops that are part of the HIV-1 Ψ-site clearly have overlapping and possibly redundant functions. SL1 has its core function in controlling the packaging of dimeric RNA. SL2 and SL3 can both

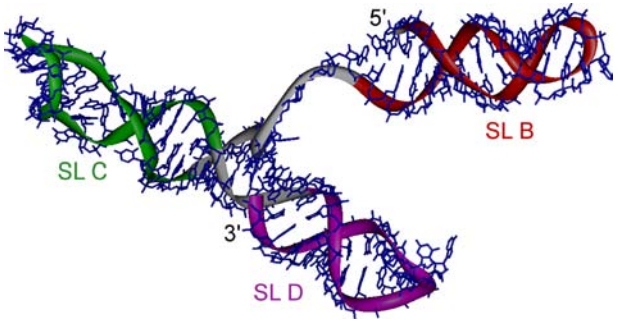


Fig.8. Structure of the core encapsidation signal RNA of MMLV (1S9S).

bind the NC protein with high affinity and specificity. SL4 only exhibits weak affinity for the NC protein and its likely function is in stabilizing the structure of the Ψ-site RNA.[72] A 19-nucleotide RNA based on the sequence of SL2 binds NC protein with an affinity (K_d = 110 nM) similar to that observed for the binding of SL3 to NC (K_d = 170 nM).[73] The solution structure of this SL2-NC complex was determined by NMR spectroscopy. In this structure the N- and C-terminal zinc knuckles of NC bind to the exposed guanine bases in the GGUG tetraloop (Fig. 9).[73] The amino acids 3–11 of the NC protein form a 3_{10}-helix that packs against the closest zinc knuckle and interacts with the RNA backbone in the stem mainly via electrostatic interactions (Fig. 9). The N-terminal zinc knuckle interacts with the A:U:A base triple platform in the minor groove of the RNA stem (Fig. 9).

Recognition of SL3 by the NC protein is achieved in a similar manner as observed for SL2. However, there are important differences in the molecular details.[73] Again, both zinc knuckles are engaged in binding to the stem-loop RNA. The N-terminal zinc knuckle binds in the major of SL2, whereas in the case of SL2 the interaction takes place with the A:U:A base triple in the minor groove. The relative orientation the N- and C-terminal knuckles differs in both complexes (Fig. 9). Sequence-specific recognition in the SL3-NC protein complex involves the guanine nucleotide bases of the loop (G7 and G9 in Fig. 9). The adenine in the loop (A8 in Fig. 9) packs against the amino-terminal zinc knuckle and forms hydrogen bonds with a conserved arginine (Arg 32). The residues 3–10 of the NC protein form a 3_{10}-helix that binds to the major groove of the SL3 stem and makes specific hydrogen bond contacts with the RNA. In contrast, the N-terminal helix in the SL2 complex is not involved in specific interactions. These differences demonstrate that the NC protein binds in an adaptive manner to SL2 and SL3 using different subsets of inter- and intramolecular interactions.

Packaging of the Dimeric Genome

Retroviruses like HIV-1 or MMLV package two copies of their genomes during the viral assembly process. This is a requirement for strand-transfer mediated recombination during reverse transcription.[74,75] The

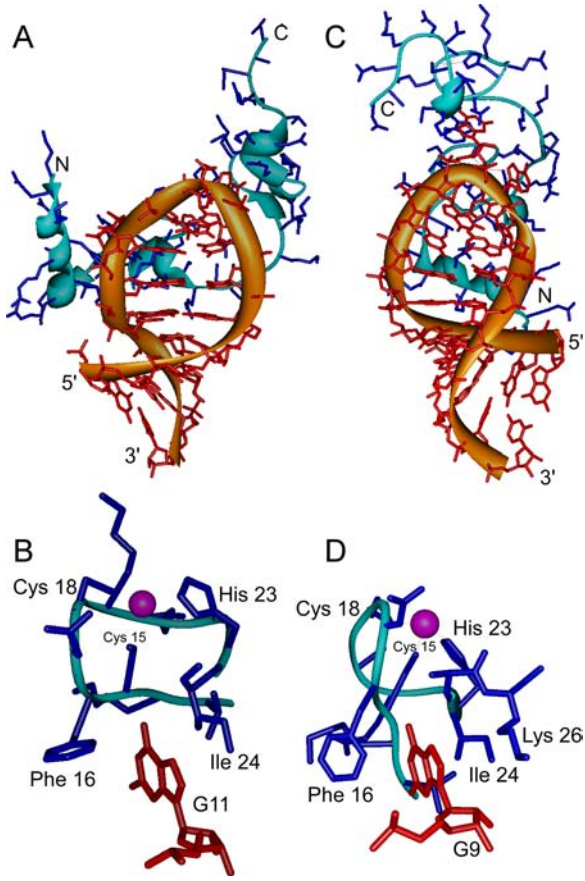


Fig. 9. Structures of HIV-1 NC protein-RNA complexes. Shown are the SL2 RNA-NC complex (A and B) and the SL3 RNA-NC complex (C and D). A and C show overviews of the two structures, B and D show the interactions of a crucial guanine nucleotide in both structures (1F6U, 1AIT).

genomic RNA exists inside the virion as a dimmer, and the overlap of the RNA sequences that control dimerization and packaging indicates that these processes are coupled.[76–78] This interlay of RNA-RNA and RNA-protein interactions was recently demonstrated for the Ψ-site RNA of MMLV, where a dimerization-induced register shift in base-pairing exposes conserved UCUG elements that bind the NC protein with very high affinity.[79] These elements are base-paired in the monomeric RNA and do not bind NC protein. This represents

essentially an RNA switch that promotes packaging of a dimeric genome by making protein binding sites in the RNA that are sequestered in the monomeric RNA available upon dimerization.

RNA-Protein Interaction in Viral RNA Packaging: Present and Future

The examples discussed above illustrate that the molecular mechanisms used by RNA viruses to ensure the incorporation of the correct genomic material are both simple and diverse. Even though the same molecular interactions — for example, stacking of bases with hydrophobic protein surfaces, or electrostatic interactions between charged amino acids and the RNA phosphate backbone — appear in all three complexes, the details in which they contribute to specific recognition differ significantly. Moreover, the same protein unit (and perhaps also the RNA) can recognize different RNA binding partners due to adaptive binding, indicating that the internal dynamics of both binding partners is crucial for the control of the interaction. The examples also demonstrate that one needs to be cautious when using secondary structures as a basis for interpreting binding data: Even though both the MS2 and the AMV packaging sites contain stem-loop structures, the loop is only involved in the recognition process during packaging of MS2. In AMV the stem-loop structures are simply providing a three-dimensional scaffold that ensures the presentation of the repeat sequences that are the key for specific recognition.

The increase in structural information about RNA packaging signals has greatly improved our understanding of the molecular details that control this crucial process in the viral lifecycle. Due to its importance for the biogenesis of infectious virus particles, we can hope that this research will in the near future provide us with a novel approach to combat viral diseases, even in cases where immunization is difficult.

Acknowledgements

The molecular graphics in this chapter were produced using the Chimera package from the Computer Graphics Laboratory, University of

California, San Francisco (supported by NIH P41 RR-01081). All coordinate files were extracted from the PDB data-base at the Rutgers Center for Structural Biology. The corresponding PDB ID numbers are given in the figure legends.

References

1. Johansson HE, Dertinger D, LeCuyer KA, *et al.* (1998) A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. *Proc Natl Acad Sci USA* **95**: 9244–9249.
2. LeCuyer KA, Behlen LS, Uhlenbeck OC. (1996) Mutagenesis of a stacking contact in the MS2 coat protein-RNA complex. *EMBO J* **15**: 6847–6853.
3. LeCuyer KA, Behlen LS, Uhlenbeck OC. (1995) Mutants of the bacteriophage MS2 coat protein that alter its cooperative binding to RNA. *Biochemistry* **34**: 10600–10606.
4. Stonehouse NJ, Scott DJ, Fonseca S, *et al.* (1996) Molecular interactions in the RNA bacteriophage MS2. *Biochem Soc Trans* **24**: 412S.
5. Valegard K, Murray JB, Stockley PG, *et al.* (1994) Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* **371**: 623–626.
6. Romaniuk PJ, Lowary P, Wu HN, *et al.* (1987) RNA binding site of R17 coat protein. *Biochemistry* **26**: 1563–1568.
7. Borer PN, Lin Y, Wang S, *et al.* (1995) Proton NMR and structural features of a 24-nucleotide RNA hairpin. *Biochemistry* **34**: 6488–6503.
8. Allain FHT, Bouvet P, Dieckmann T, Feigon J. (2000) Molecular basis of sequence specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J* **19**: 6870–6881.
9. Bouvet P, Finger DL, Allain FHT, *et al.* (2001) Recognition of preformed and flexible elements of an RNA stem-loop by nucleolin. *J Mol Biol* **309**: 763–775.
10. Finger LD, Trantirek L, Johansson C, Feigon J. (2003) Solution structures of stem-loop RNAs that bind to the two N-terminal RNA-binding domains of nucleolin. *Nucleic Acids Research* **31**: 6461–6472.
11. Hall KB. (1994) Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry* **33**: 10076–10088.
12. Valegard K, Murray JB, Stonehouse NJ, *et al.* (1997) The three-dimensional structures of two complexes between recombinant MS2

capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J Mol Biol* **270**: 724–738.
13. Grahn E, Moss T, Helgstrand C, *et al.* (2001) Structural basis of pyrimidine specificity in the MS2 RNA hairpin-coat-protein complex. *RNA* **7**: 1616–1627.
14. Grahn E, Stonehouse NJ, Murray JB, *et al.* (1999) Crystallographic studies of RNA hairpins in complexes with recombinant MS2 capsids: implications for binding requirements. *RNA* **5**: 131–138.
15. Helgstrand C, Grahn E, Moss T, *et al.* (2002) Investigating the structural basis of purine specificity in the structures of MS2 coat protein RNA translational operator hairpins. *Nucleic Acids Res* **30**: 2678–2685.
16. van den Worm SH, Stonehouse NJ, Valegard K, *et al.* (1998) Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments. *Nucleic Acids Res* **26**: 1345–1351.
17. Baidya N, Uhlenbeck OC. (1995) The role of 2′-hydroxyl groups in an RNA-protein interaction. *Biochemistry* **34**: 12363–12368.
18. Grahn E, Stonehouse NJ, Adams CJ, *et al.* (2000) Deletion of a single hydrogen bonding atom from the MS2 RNA operator leads to dramatic rearrangements at the RNA-coat protein interface. *Nucleic Acids Res* **28**: 4611–4616.
19. Shtatland T, Gill SC, Javornik BE, *et al.* (2000) Interactions of Escherichia coli RNA with bacteriophage MS2 coat protein: genomic SELEX. *Nucleic Acids Res* **28**: E93.
20. Stockley PG, Stonehouse NJ, Murray JB, *et al.* (1995) Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. *Nucleic Acids Res* **23**: 2512–2518.
21. Munshi S, Liljas L, Cavarelli J, *et al.* (1996) The 2.8 A structure of a T = 4 animal virus and its implications for membrane translocation of RNA. *J Mol Biol* **261**: 1–10.
22. Witherell GW, Gott JM, Uhlenbeck OC. (1991) Specific interaction between RNA phage coat proteins and RNA. *Prog Nucleic Acid Res Mol Biol* **40**: 185–220.
23. Mattaj IW, Nagai K. (1995) Recruiting proteins to the RNA world. *Nat Struct Biol* **2**: 518–522.
24. Bol JF. (1999) Alfalfa mosaic virus and ilarviruses: involvement of coat protein in multiple steps of the replication cycle. *J Gen Virol* **80** (Pt 5): 1089–1102.
25. Quadt R, Rosdorff HJ, Hunt TW, Jaspars EM. (1991) Analysis of the protein composition of alfalfa mosaic virus RNA-dependent RNA polymerase. *Virology* **182**: 309–315.

26. Jaspars EM. (1985) Interaction of alfalfa mosaic virus nucleic acid and protein. In JW Davies (ed), *Molecular Plant Virology*, pp. 155–221. CRC Press, New York.
27. Ansel-McKinney P, Scott SW, Swanson M, *et al.* (1996) A plant viral coat protein RNA binding consensus sequence contains a crucial arginine. *EMBO J* **15**: 5077–5084.
28. Houser-Scott F, Baer ML, Liem Jr KF, *et al.* (1994) Nucleotide sequence and structural determinants of specific binding of coat protein or coat protein peptides to the 3′ untranslated region of alfalfa mosaic virus RNA 4. *J Virol* **68**: 2194–2205.
29. Swanson MM, Ansel-McKinney P, Houser-Scott F, *et al.* (1998) Viral coat protein peptides with limited sequence homology bind similar domains of alfalfa mosaic virus and tobacco streak virus RNAs. *J Virol* **72**: 3227–3234.
30. Yusibov V, Loesch-Fries LS. (1998) Functional significance of three basic N-terminal amino acids of alfalfa mosaic virus coat protein. *Virology* **242**: 1–5.
31. Yusibov VM, Loesch-Fries LS. (1995) N-terminal basic amino acids of alfalfa mosaic virus coat protein involved in the initiation of infection. *Virology* **208**: 405–407.
32. Baer ML, Houser F, Loesch-Fries LS, Gehrke L. (1994) Specific RNA binding by amino-terminal peptides of alfalfa mosaic virus coat protein. *EMBO J* **13**: 727–735.
33. Fukuyama K, Abdel-Meguid SS, Johnson JE, Rossmann MG. (1983) Structure of a T = 1 aggregate of alfalfa mosaic virus coat protein seen at 4.5 A resolution. *J Mol Biol* **167**: 873–890.
34. Kumar A, Reddy VS, Yusibov V, *et al.* (1997) The structure of alfalfa mosaic virus capsid protein assembled as a T = 1 icosahedral particle at 4.0 A resolution. *J Virol* **71**: 7911–7916.
35. Guogas LM, Filman DJ, Hogle JM, Gehrke L. (2004) Cofolding organizes alfalfa mosaic virus RNA and coat protein for replication. *Science* **306**: 2108–2111.
36. Coffin JM, Hughes SH, Varmus HE. (1997) *Retroviruses.* Cold Spring Harbor Laboratory Press, New York.
37. Demene H, Dong CZ, Ottmann M, *et al.* (1994) H-1 NMR structure and biological studies of the His(23)-]Cys mutant nucleocapsid protein of HIV-1 indicate that the conformation of the first zinc finger is critical for virus infectivity. *Biochemistry* **33**: 11707–11716.
38. Lee BM, De Guzman RN, Turner BG, *et al.* (1998) Dynamical behavior of the HIV-1 nucleocapsid protein. *J Mol Biol* **279**: 633–649.

39. Omichinski JG, Clore GM, Sakaguchi K, *et al.* (1991) Structural characterization of a 39-residue synthetic peptide containing the two zinc binding domains from the HIV-1 p7 nucleocapsid protein by CD and NMR spectroscopy. *FEBS Lett* **292**: 25–30.
40. South TL, Blake PR, Sowder RC, *et al.* (1990) The nucleocapsid protein isolated from HIV-1 particles binds zinc and forms retroviral-type zinc fingers. *Biochemistry* **29**: 7786–7789.
41. Summers MF. (1991) Zinc finger motif for single-stranded nucleic acids? Investigations by nuclear magnetic resonance. *J Cell Biochem* **45**: 41–48.
42. Summers MF, South TL, Kim B, Hare DR. (1990) High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach. *Biochemistry* **29**: 329–340.
43. Demene H, Jullian N, Morellet N, *et al.* (1994) Three-dimensional 1H NMR structure of the nucleocapsid protein NCp10 of Moloney murine leukemia virus. *J Biomol NMR* **4**: 153–170.
44. Klein DJ, Johnson PE, Zollars ES, *et al.* (2000) The NMR structure of the nucleocapsid protein from the mouse mammary tumor virus reveals unusual folding of the C-terminal zinc knuckle. *Biochemistry* **39**: 1604–1612.
45. Gao Y, Kaluarachchi K, Giedroc DP. (1998) Solution structure and backbone dynamics of Mason-Pfizer monkey virus (MPMV) nucleocapsid protein. *Protein Sci* **7**: 2265–2280.
46. Aldovini A, Young RA. (1990) Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J Virol* **64**: 1920–1926.
47. Dupraz P, Oertle S, Meric C, *et al.* (1990) Point mutations in the proximal Cys-His box of Rous sarcoma virus nucleocapsid protein. *J Virol* **64**: 4978–4987.
48. Gorelick RJ, Henderson LE, Hanser JP, Rein A. (1988) Point mutants of moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a "zinc finger-like" protein sequence. *Proc Natl Acad Sci USA* **85**: 8420–8424.
49. Meric C, Gouilloud E, Spahr PF. (1988) Mutations in Rous sarcoma virus nucleocapsid protein p12 (NC): deletions of Cys-His boxes. *J Virol* **62**: 3328–3333.
50. Sakalian M, Wills JW, Vogt VM. (1994) Efficiency and selectivity of RNA packaging by Rous sarcoma virus Gag deletion mutants. *J Virol* **68**: 5969–5981.

51. Meric C, Goff SP. (1989) Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein. *J Virol* **63**: 1558–1568.
52. Dannull J, Surovoy A, Jung G, Moelling K. (1994) Specific binding of HIV-1 nucleocapsid protein to PSI RNA *in vitro* requires N-terminal zinc finger and flanking basic amino acid residues. *EMBO J* **13**: 1525–1533.
53. Gorelick RJ, Chabot DJ, Rein A, *et al.* (1993) The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent. *J Virol* **67**: 4027–4036.
54. Tanchou V, Decimo D, Pechoux C, *et al.* (1998) Role of the N-terminal zinc finger of human immunodeficiency virus type 1 nucleocapsid protein in virus structure and replication. *J Virol* **72**: 4442–4447.
55. Clavel F, Orenstein JM. (1990) A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J Virol* **64**: 5230–5234.
56. Poznansky M, Lever A, Bergeron L, *et al.* (1991) Gene transfer into human lymphocytes by a defective human immunodeficiency virus type 1 vector. *J Virol* **65**: 532–536.
57. Berkowitz RD, Fisher J, Goff SP. (1996) RNA packaging. *Curr Top Microbiol Immunol* **214**: 177–218.
58. Baudin F, Marquet R, Isel C, *et al.* (1993) Functional sites in the 5′ region of human-immunodeficiency-virus type-1 RNA form defined structural domains. *J Mol Biol* **229**: 382–397.
59. Berkowitz RD, Goff SP. (1994) Analysis of binding-elements in the human-immunodeficiency-virus type-1 genomic RNA and nucleocapsid protein. *Virology* **202**: 233–246.
60. Berkowitz RD, Luban J, Goff SP. (1993) Specific binding of human-immunodeficiency-virus type-1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J Virol* **67**: 7190–7200.
61. Clever J, Sassetti C, Parslow TG. (1995) RNA secondary structure and binding-sites for gag gene-products in the 5′-packaging signal of human-immunodeficiency-virus type-1. *J Virol* **69**: 2101–2109.
62. Clever JL, Parslow TG. (1997) Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol* **71**: 3407–3414.
63. McBride MS, Panganiban AT. (1996) The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J Virol* **70**: 2963–2973.

64. McBride MS, Panganiban AT. (1997) Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation *in vivo*. *J Virol* **71**: 2050–2058.
65. Sakaguchi K, Zambrano N, Baldwin ET, *et al.* (1993) Identification of a binding-site for the human-immunodeficiency-virus type-1 nucleocapsid protein. *Proc Natl Acad Sci USA* **90**: 5219–5223.
66. Turner BG, Summers MF. (1999) Structural biology of HIV. *J Mol Biol* **285**: 1–32.
67. Ennifar E, Yusupov M, Walter P, *et al.* (1999) The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure* **7**: 1439–1449.
68. Mujeeb A, Clever JL, Billeci TM, *et al.* (1998) Structure of the dimer initiation complex of HIV-1 genomic RNA. **5**: 432–436.
69. Lawrence DC, Stover CC, Noznitsky J, *et al.* (2003) Structure of the intact stem and bulge of HIV-1 psi-RNA stem-loop SL1. *J Mol Biol* **326**: 529–542.
70. Amarasinghe GK, De Guzman RN, Turner RB, Summers MF. (2000) NMR structure of stem-loop SL2 of the HIV-1 Psi RNA packaging signal reveals a novel A-U-A base-triple platform. *J Mol Biol* **299**: 145–156.
71. D'Souza V, Dey A, Habib D, Summers MF. (2004) NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J Mol Biol* **337**: 427–442.
72. Amarasinghe GK, Zhou J, Miskimon M, *et al.* (2001) Stem-loop SL4 of the HIV-1 Psi-RNA packaging signal exhibits weak affinity for the nucleocapsid protein. Structural studies and implications for genome recognition. *J Mol Biol* **314**: 961–970.
73. Amarasinghe GK, De Guzman RN, Turner RB, *et al.* (2000) NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Psi-RNA packaging signal. Implications for genome recognition. *J Mol Biol* **301**: 491–511.
74. Hu WS, Temin HM. (1990) Retroviral recombination and reverse transcription. *Science* **250**: 1227–1233.
75. Varmus HE. (1982) Form and function of retroviral proviruses. *Science* **216**: 812–820.
76. Housset V, De Rocquigny H, Roques BP, Darlix JX. (1993) Basic amino acids flanking the zinc finger of Moloney murine leukemia virus

nucleocapsid protein NCp10 are critical for virus infectivity. *J Virol* **67**: 2537–2545.

77. Paillart JC, Shehu-Xhilaga M, Marquet R, Mak J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol* **2**: 461–472.
78. Prats AC, Roy C, Wang PA, *et al.* (1990) Cis elements and trans-acting factors involved in dimer formation of murine leukemia virus RNA. *J Virol* **64**: 774–783.
79. D'Souza V, Summers MF. (2004) Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature* **431**: 586–590.

This page intentionally left blank

Chapter 9

Rational Design of Viral Protein Structures with Predetermined Immunological Properties

James Lara[†] *and Yury Khudyakov*[*,†]

Viral pathogens often stimulate the infected host to develop an immunologic response to viral proteins as protection against infection. Researchers have explored this human immune response in efforts to develop vaccines against viral infections, immunotherapeutics for the treatment of diseases caused by these infections, and diagnostic reagents for the detection of current or past infections. The breadth of applications for viral proteins in all three domains of medicine — diagnostics, prophylaxis, and therapy — has spurred a significant demand for these products.

The three major sources of viral proteins for medical applications are: 1) naturally infected persons, 2) *in vitro*-infected tissues and cell cultures, and 3) recombinant DNA technology. Historically, in the absence of any other reliable source, tissues from infected persons have been used to support the development of vaccines and diagnostics.[1–11] One of the most well-known examples is the development of hepatitis B vaccine using hepatitis B virus (HBV) surface antigen particles derived from plasma from rejected blood donations.[9,12,13] Notable uses

*Corresponding author.
[†]Division of Viral Hepatitis, Centers for Disease Control and Prevention, 1600 Clifton Road MS-A33, Atlanta, Georgia, 30333, USA.

of antigens derived from *in vitro* sources include influenza A vaccine derived from *in vitro*-infected embryonated chicken eggs[14,15] and hepatitis A virus (HAV) vaccine and diagnostic reagents obtained from *in vitro*-infected cell cultures.[16–18] However, despite the number of vaccines and diagnostic materials developed from *in vivo* or *in vitro* sources and their use in the prevention and diagnosis of many common viral infections, serious concerns exist about the safety and cost-effectiveness of viral materials derived from tissues infected with live virus.[19–26] Mistakes in the handling of infectious material prepared for vaccine production can have disastrous consequences. The distribution of infectious virus with some batches of the original Sabin poliovirus vaccine in the early days of a large-scale production of this vaccine is one of the most infamous accidents in the recent history of live vaccines.[27]

The application of recombinant DNA technology to medicine promised an immediate and dramatic improvement in the safety of inexpensive preparations of viral proteins. A notable example is the use of HBV surface antigen obtained from transformed yeast to create an efficient, safe, and affordable recombinant vaccine.[2,8,28–33] However, the application of recombinant DNA technology to hepatitis B vaccine development is unlikely to be duplicated for other viruses such as human immunodeficiency virus (HIV) and hepatitis C virus (HCV), which tenaciously defy attempts to develop an efficient preventive vaccine. To a significant degree, the difficulties in developing HIV and HCV vaccines are related to the limited knowledge of neutralizing immunoresponses against different strains of these viruses.[34–40]

As researchers realized that the straightforward approach to development of a recombinant hepatitis B vaccine could not be applied to other viruses and that new strategies were needed, they turned to protein engineering, which was born in the early 1980s. Protein engineering is the application of molecular techniques and scientific principles to the design and construction of novel proteins with desired properties. The spectacular developments over the last three decades in the automatic high-throughput chemical synthesis of long peptides[41–43] and deoxypolynucleotides[44] have built a solid technological foundation for protein engineering. Although these advances in molecular technology made the construction of very large proteins of

any amino acid sequence technically feasible, the currently limited understanding of protein structure and of the relationship between structure and function has hindered the application of protein engineering to many complex medical problems.

Rational design of proteins is rooted in the knowledge of the chemical, physical, and biological properties of proteins. In those cases when sufficient knowledge of these properties was secured, a significant progress in protein engineering was attained.[45] On the other hand, recent progress in the assembly, maintenance, and screening of large libraries of peptides and proteins and in computational approaches to protein analysis and design caused an explosive development of new strategies in engineering of proteins with predetermined immunological properties. Some of these strategies can be used for the improvement of immunological activity of proteins and some for the *de novo* synthesis of proteins with new immunological properties.

This chapter reviews major strategies for engineering of viral proteins with predetermined immunological properties. These strategies can be classified into "structural," "functional," and "focused" approaches. The structural approach is as close to the real rational design of proteins as the modern state of science allows. However, because detailed knowledge of a protein's quantitative structure-activity relationship (QSAR) is often unavailable, this approach results more often in devastating frustrations than in exalting victories, which was the main reason for the development of the functional approach. Also known as "directed evolution," the functional approach requires no prior knowledge of a protein's QSAR. Rather, it is based on the availability of a representative library of peptides or proteins and a selection procedure for the peptide or protein with the desired activity. It is termed a "functional" approach because proteins are selected based on specific functions without regard to structure. The third strategy, referred to here as the focused approach, avoids the limitations of the other two approaches by applying mathematical modeling to gain new QSAR knowledge and use this knowledge to design proteins with improved immunological properties. The first two approaches will be reviewed only briefly below. The chapter highlights the focused approach.

Strategies for Protein Design

Structural protein design. The structural approach is the cradle of protein engineering, marked by a history of more than two decades of audacious attempts to produce proteins with improved biological properties. Although not always successful, these engineering experiments thoroughly tested the limits of current QSAR knowledge and helped identify areas requiring further research. In the medical field, the attempts centered mainly on the development of new approaches to vaccines and diagnostic reagents.

The origin of the structural approach can be traced to the first attempt to model viral antigenic properties with short synthetic peptides. The strategies for chemical synthesis of peptides were developed at the dawn of the 20th century by chemist Emil Fischer. However, it was only after 1963, when Bruce Merrifield pioneered his method of solid-phase peptide synthesis (for which he was awarded the Nobel Prize almost 20 years later), that the real applications of peptides were developed in different fields of protein research. In 1963, Anderer first observed that a short peptide could elicit antibodies immunoreactive with the intact virus.[46] These two important advances — namely, solid-phase peptide synthesis and modeling antigenic epitopes with short peptides — opened the door for a frontal attack on the nature of viral antigenic epitopes and the application of this new knowledge to the design of vaccines and diagnostics. It was not until more than a decade later, however, that peptide research yielded important discoveries in this area and the field made an important next stride toward this goal.

In 1976, Michael Sela, who studied the neutralizing immunoresponse to MS2 phage, discovered that a short synthetic peptide, if properly presented to immunocompetent cells, could elicit a neutralizing immunoresponse. This research prompted other studies using synthetic peptides, which by that time had become available to researchers in sufficient quantities. Synthetic peptides as short as five amino acids were found to bind antibodies elicited against a whole protein (from which the peptides were derived) and to elicit antibodies that specifically recognize this protein.[47]

The initial success of the research with synthetic peptides presented an attractive opportunity to develop theoretical approaches for the prediction of B- and T-cell-specific antigenic epitopes within proteins[48,49] so that these epitopes could be used to design proteins with desired immunological properties. Although not very accurate, the computational approaches were used extensively by many laboratories to identify antigenic sites in proteins. Subsequently, development of automated methods for peptide synthesis allowed epitope identification through total scanning of proteins with overlapping synthetic peptides.[50,51] Although often efficient, this "brute force" approach to protein scanning was wasteful and limited to identification of continuous linear antigenic epitopes. It was also inadequate for identification of conformation-dependent "split" epitopes, which are often involved in virus neutralization.

Recognizing the need for more accurate computational approaches to epitope prediction, many laboratories undertook research efforts associated with immunological studies. Subsequently, advances in the identification of mechanisms of antigen presentation and recognition drastically accelerated progress in the development of sophisticated computational techniques for the prediction of T-cell epitopes.[48,49] Unfortunately, computational approaches to B-cell epitope prediction did not receive proper development and remain either simplistic or dependent on a comprehensive knowledge of protein structure,[52,53] which is rarely available. Despite the limited efficacy of these approaches, many antigenic epitopes were nevertheless identified and used in protein engineering experiments. It was discovered that short antigenic epitopes, which can be modeled with synthetic peptides, can usually be reliably reproduced in a functional form when inserted into recombinant proteins.[54–59] This important feature made antigenic epitopes attractive building blocks for the "structural" design of proteins with predetermined immunological properties.

Two approaches to structural design using short antigenic epitopes to construct immunologically active proteins have been devised for the development of new antigenic targets for serologic assays and vaccine candidates. One approach, based on construction of completely artificial antigens (multiple epitope proteins [MEP]) built of

short antigenic regions, is used mainly to develop reagents with improved diagnostically relevant properties. Because antibody binding is a relatively simple protein-protein interaction process and/or because antigenic epitopes are usually discovered by binding with antibodies directly (the most essential property for diagnostic targets), the MEP approach can be used to obtain artificial antigens suitable for diagnostics.[60–64] The structural design of vaccines is more difficult because antibody eliciting is a significantly more complex biological process and the pathogen neutralization mechanisms are not fully understood. In this case, antigenic epitopes of interest are frequently inserted into carrier proteins that serve to stimulate an immunoresponse against the inserted epitopes.

The MEP strategy involves the use of broadly and strongly immunoreactive antigenic epitopes to design and construct synthetic genes that encode artificial polypeptides composed of these epitopes. MEPs imitate only the diagnostically relevant immunologic properties of natural antigens without regard for other associated functions. Because MEPs are composed of epitopes that are relevant only to diagnostics, they are expected to significantly reduce the opportunity for nonspecific reactivity[65–70] since ~3%–4% of pathogen-specific antibodies may cross-immunoreact with host-specific antigens.[69] The MEP strategy is flexible and allows researchers to combine in one polypeptide several copies of antigenic epitopes from many different antigens and sequence variants of these antigens.[61–64]

Despite its advantages, the MEP strategy has not become a mainstream method for obtaining diagnostic reagents. Since the strategy relies on the use of antigenic epitopes modeled with synthetic peptides, it does not involve conformational epitopes, which comprise a significant pool of diagnostically relevant epitopes. In addition, combining antigenic regions in one protein generates many junctions between these regions, which could potentially generate new epitopes and interfere with accurate antibody detection. Although antigenic epitopes are reproduced when inserted into artificial antigen, there is no assurance that these epitopes are fully functional. Artificial combination of several epitopes also raises concerns about adverse interference between these epitopes, which may render some epitopes to be

less active or even inaccessible for antibody binding. Finally, important protein properties such as solubility and stability, which are difficult to predict, may make the artificial antigen unsuitable for diagnostics. Thus, limited QSAR knowledge does not completely preclude the engineering of artificial proteins with some useful diagnostic properties, but nevertheless does not allow for the reliable rational design of superior diagnostic targets.

An alternative strategy based on the use of carrier proteins is applied mainly to engineering of vaccines. A carrier protein is usually selected based on its immunogenicity or capacity to induce a strong immunoresponse. If a carrier protein is highly immunogenic, the design of an artificial protein capable of eliciting an immunoresponse to some desired antigenic epitopes involves only inserting those epitopes into the carrier protein so that the inserts do not adversely affect the immunogenicity of the carrier and would be recognized by immunocompetent cells responsible for producing the antibody response. Although many proteins can be used as carrier proteins, the nucleocapsid, capsid, and envelope proteins of viruses are the most useful. These proteins expressed in heterologous systems usually maintain their propensity to assemble into complex aggregates with regular structures reminiscent of virions. The recombinant protein structures are frequently referred to as virus-like particles (VLPs). Many researchers consider VLPs , which seem to be assembled into "natural" conformations, as the most suitable protein structures for efficient antigen presentation and/or clonal selection during the development of an immunoresponse to viruses.[45] Because of their strong antigenic and immunogenic reactivity and their use as targets for protein engineering, VLPs have been extensively explored as diagnostic reagents and vaccines.[71–73]

Although many laboratories have devoted significant effort to the development of the VLP strategy and have obtained encouraging results, the approach remains primarily in the domain of research and has yet to be applied to actual vaccine development. The major problem is that the strategy imitates active protein structures that have several important functions (e.g. packaging and protection of viral genetic material, cell receptor binding, cell penetration, and regulatory

interactions with host proteins) among which immunological activities are of somewhat peripheral importance. Since the determinants of high immunogenicity in these protein structures are unknown, researchers are forced to imitate the structures in their entirety in the hope that the desired immunological properties will be reproduced. However, even though immunological properties are modeled with these structures, many other properties are almost certainly reproduced as well. In addition, VLPs are complex protein structures that contain many antigenic epitopes, only a fraction of which contribute to the intended use in virus neutralization. All of these factors limit the value of this strategy for vaccine engineering.

Although all currently available vaccines are affected by similar constraints, rational vaccine design implies improvement, rather than reproduction, of protein properties by facilitating the desired immunoresponse and excluding or at least reducing any potential shortcomings. Rationally designed proteins should perform to specifications and should not include functions unrelated to their primary purpose. The design of artificial antigens or immunogens should therefore be limited to relevant immunological properties without regard to other properties associated with the naturally occurring antigens. Such design can be achieved using building blocks with known simple elementary functions. However, since QSAR for virus neutralization is not sufficiently understood, such elementary building blocks are not available. Therefore, vaccines cannot yet be constructed using this design. This lack of knowledge limits engineering choices to the use of rather complex structures such as VLPs to stimulate desired immunoresponse at the price of including many properties unrelated to pathogen neutralization.

Researchers can overcome the limitations of the carrier protein strategy to some extent by using a carrier derived from the same virus for which the vaccine is engineered. For example, use of the HBV nucleocapsid protein (C), which is efficiently assembled into VLP in many expression systems,[74–82] as a carrier protein for the neutralizing epitopes derived from HBV surface antigen (HBsAg) allows the neutralizing properties of both proteins to be combined.[77,79,83–86] To reduce irrelevant immunological activity, VLPs that elicit a protective

immunoresponse against one virus can be used as carrier proteins for the neutralizing epitopes derived from the other virus. For example, HBsAg can be used for the insertion of antigenic epitopes from HIV, thus providing a multivalent construct that elicits antibodies against both viruses.[87,88]

As noted above, the real protein engineering begins when simple building blocks with clearly known structure-function connections are available for the construction of structures with desired properties. The carrier protein strategy uses building blocks that are relatively simple but not completely understood. In many cases, the approach requires a trial of several constructs or expression systems because of limited understanding of the factors deemed important for vaccine engineering, such as aggregation, structure stability, and stimulation of an immunoresponse against the insert.[45]

Functional strategy. Application of the structural approach to protein design is constrained by a limited understanding of the QSAR for immunological properties of viral proteins. To address this limitation, researchers devised the "functional" strategy for obtaining proteins with desired functions in the absence of structural knowledge. The basis of this strategy is simple and almost as old as humankind, i.e. nature's method of evolution. Throughout history, humans have used a similar approach to breed domestic animals and cultivate plants. Applied to proteins, the approach became known as directed evolution (DE), or molecular breeding. Although the DE approach does not require any degree of rational design, it is helpful in obtaining new proteins and proteins with improved properties, and it is currently the protein engineering method of choice in many laboratories.

The DE approach involves the development of a library of protein sequence variants and a procedure for selecting variants with desired immunological properties (e.g. ability to bind with antibodies). If a library includes an exhaustive array of sequence variants, then the DE approach theoretically allows for the selection of all protein structures that, for example, can bind a particular monoclonal antibody. Such an exhaustive library can be obtained using random protein sequences. In the example of monoclonal antibody binding, the power of this strategy is in the opportunity to identify not only naturally occurring

structures but also new structures that may bind this monoclonal antibody. The most amazing feature of the approach is that all findings can be attained without any prior QSAR knowledge.[89–92] The specifically active protein structures are, however, discovered rather than designed.

Peptide variants selected from the random sequence peptide library by use of antibody binding frequently do not reproduce an exact copy, but rather mimic natural antigenic epitopes.[93–95] The imitated epitopes are known as mimotopes. The strategy for obtaining mimotopes has several important applications, including epitope mapping within proteins.[90,96] However, the most valuable application of this strategy is in obtaining short sequences to model conformational antigenic epitopes that can otherwise be reproduced only by using long stretches of protein chains.[92–94] The limitation of the strategy is its feasibility only for very short peptide sequences. A comprehensive library of 20-mer peptides, in which each sequence variant has a chance to be represented at least once, should contain >400 kg of peptides. Thus, sequences in the range of 100 aa are completely out of the scope of this approach, and the method is usually applied to peptides <10 aa in length.[92,97]

The mimotope strategy can also be used to improve natural epitopes in terms of characteristics such as antibody-binding affinity.[98] However, a more efficient way to accomplish this task is to use a set of closely related sequence variants of an antigenic region for the selection of the improved variant.[99] This set can be obtained by randomization of the original sequence. The randomization can be achieved, for example, by polymerase chain reaction using a low-fidelity DNA polymerase for the amplification of a DNA fragment encoding for the protein of interest or a region of this protein.[100,101] A significant improvement in the protein properties can be attained by applying several cycles of randomization and selection.[101]

Several techniques can be used to implement the DE strategy. Some techniques use recombinant DNA, whereas others use synthetic peptides. Phage display, one of the most popular recombinant DNA techniques, is based on the expression of a protein or its fragments onto the surface of filamentous phages, the genome of which contains genes encoding these proteins.[102] This versatile technique was used

for the identification of mimotopes[92] and the optimization of antigenic properties for large antigens.[103] A similar concept of antigen presentation was used for the development of cell and ribosome display techniques.[104] Another popular technology, based on the use of libraries of synthetic peptides with random sequences,[105,106] is applicable only to identification of short peptide ligands and mimotopes.

The DE techniques mentioned above use heterogeneity on the level of individual amino acid positions. This feature makes these techniques both comprehensive and limited in application. A technique known as DNA shuffling takes advantage of recombination — another source of heterogeneity in nature — for the generation of protein heterogeneity, which can be used for the selection of recombinants with improved immunological properties.[107,108] This technique achieves rapid evolution of genes under *in vitro* selection pressure through an iterative process of recombination and selection and has been used to develop new vaccine candidates against pathogenic agents including dengue virus, Venezuelan equine encephalitis virus, HIV, and HBV.[109]

Despite its advantages, directed evolution is rather limited in its applications. The approach does not build QSAR knowledge and lacks rational design. Because the knowledge base does not grow, building a new protein requires repeating the entire DE process from the beginning. Although the structural approach to protein engineering also makes no direct contributions to QSAR knowledge, it challenges existing concepts and identifies research gaps. The DE approach, on the other hand, is used as a "magic wand" that may or may not bring desired results. Because the approach does not always yield proteins with required specifications, it has limited applicability for building new proteins. Using the DE strategy, new protein structures can be obtained only for rather simple properties such as protein-ligand or protein-protein interactions that can be easily modeled with very short peptides. More complex properties, such as enzymatic activities, require scanning a much greater protein sequence space that cannot be accessed using random sequence peptide libraries of a manageable size. Thus, directed evolution is not protein engineering *per se*, but is more aptly characterized as molecular breeding.

"Focused" strategy. The focused approach to protein engineering overcomes many of the limitations of the other two approaches. The structural approach, which is based on global QSAR knowledge, is inefficient because of the shortage of such knowledge. The functional approach, which is based on *in vitro* directed evolution of proteins, bypasses rational design but has limited capacity to build proteins to functional specifications. Although still at its inception and limited in application, the focused approach promises significant progress in protein engineering due to its capacity for direct generation of the QSAR knowledge required for the rational design of proteins with predetermined properties. Whereas the structural strategy reveals gaps in knowledge and the functional approach "breeds" desired properties without building new knowledge, the focused strategy breeds new knowledge that is focused on obtaining new protein structures with desired properties.

At the core of the focused approach is computer-assisted mathematical modeling of the association between the structural and functional properties of proteins. The mathematical model can be built using a rather limited data set of the protein structural variants displaying variations in its activity. Although the size of the data set affects the accuracy of the model, the model can be applied as a guide for the virtual exploration of sequence space in search of protein structures with improved or new properties. The new structures found through this virtual process can be constructed and tested experimentally in the laboratory. If the desired properties are not matched by the predicted proteins, indicating inaccurate modeling, the new data obtained from the evaluation can be used to adjust the model and the mathematical modeling process. The virtual search and experimental testing can be repeated until satisfactory results are obtained. Thus, the focused strategy is a cyclic process designed to accelerate the development or evolution of knowledge by use of a mathematical model. The major advantage of this approach is its use of a virtual process instead of experimentation to evaluate structural variants for a desired function. Although the functional approach can also screen vast protein sequence spaces using experimental approaches, it relies on preexisting protein libraries and is impractical

for use with proteins >10–15 aa in length. In contrast, the focused strategy can start with a limited number of protein variants and still achieve the desired properties through a guided virtual search among an astronomical number of protein sequence variants.

At present, the focused strategy is rather limited and complicated. It cannot be used to build proteins without sufficient knowledge, i.e. a reliable mathematical model. Although application of the focused strategy results in identification of the desired protein, it is not true protein design or engineering but rather a virtual search or protein sequence space exploration for a protein structure with satisfactory properties. However, in contrast to the other two strategies discussed above, this approach builds QSAR knowledge and therefore directly facilitates development of elementary protein building blocks. Thus, the focused strategy has two important roles: 1) to build new proteins with predetermined properties by focusing mathematical modeling on the protein engineering task, and 2) to build a solid foundation for the efficient application of the structural strategy by generating elementary building blocks for protein engineering.

Focused Strategy: Mathematical Modeling

The focused strategy is combinatorial in nature and requires an iterative process that begins with any viral molecule that displays a desired biological profile and ends with an optimization of the activity profile for that molecule. The process requires sophisticated computational techniques to efficiently sample the sequence space for a given application. The combinatorial possibilities of this strategy for even simple systems are explosive and cannot be handled experimentally. For example, the number of molecules required to interchange 20 amino acids within the 4-aa peptide is 20^4, corresponding to 160,000 possible peptides. The alternative to the experimental approach to protein optimization is development of a theory, i.e. a mathematical model that quantitatively relates variations in biological activity to changes in molecular descriptors (e.g. physicochemical properties of amino acids) that can easily be obtained for each protein. One mechanism to reduce the search space is to restrict the types of molecules generated

to those that meet the requirements of the application area. Defining appropriate constraints and ensuring that the molecules generated are consistent with the constraints becomes very important. Focusing the design according to physicochemical constraints can be an effective way to prevent inappropriate molecule candidates, guide the search in space, and derive a QSAR that determines the immunological properties of proteins.

The process begins with the conception of a mathematical model that relates the chemical features of the molecule (or series of molecules) to a biological activity, such as immunogenicity or antigenicity. Without any detailed understanding of the process(es) responsible for the activity, the model is refined by examining structural similarities and differences for molecules with variations in immunological properties. Molecules are selected to maximize the presence of functional groups and structural features believed to be responsible for the desired activity. Such a strategy can provide the knowledge and computational capability for *de novo* design and engineering of protein molecules with desired immunological properties. The following is a discussion of some of the computational and mathematical methods used to derive QSAR. These include artificial neural networks, evolutionary algorithms, and Bayesian networks.

Artificial Neural Networks

Artificial neuron model. The artificial neuron model is a fairly new form of artificial intelligence (AI) based on brain theory.[110–112] An artificial neural network (ANN) is a mathematical model of neuron operation that can, in principle, be used to compute any arithmetic or logical function. It consists of two types of components or elements: 1) processing, or computation, units (also called "neurons"); and 2) connections, with adjustable "weights" or "strengths." Figure 1(A) shows the model of a single neuron (the basis of an ANN) and the two mathematical terms that describe an artificial neuron k. The fundamental operation of an artificial neuron is carried out in three steps. First, a signal x_1 at the input of synapse j connected to neuron k is multiplied by the synaptic weight w_{kj}. The synaptic weight of an artificial

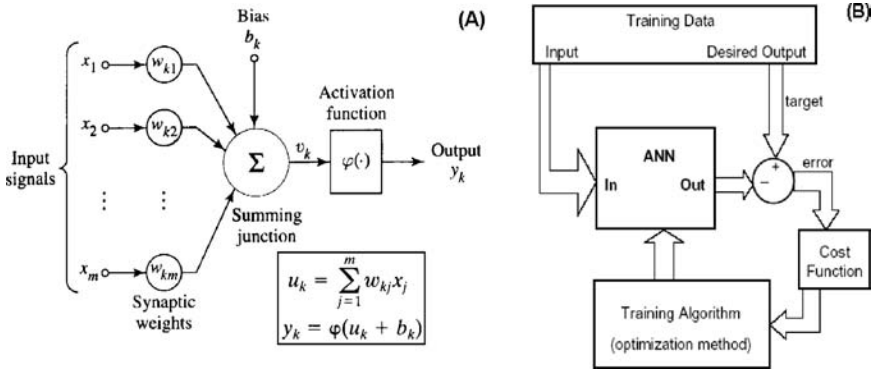


Fig 1. (**A**) ANN "single preceptron" model. Inbox shows the two mathematical terms to describe a neuron k, where x_1, x_2, … , x_m are the input signals (e.g. amino acids in the sequence); w_{k1}, w_{k2}, … , w_{km} are the synaptic weights of neuron k; u_k is the linear combiner output due to the input signals; b_k is the bias; $\varphi(\bullet)$ is the activation function; and y_k is the output signal of the neuron; the use of bias b_k has the effect of applying an affine transformation to the output u_k of the linear combiner in the model, as shown by $v_k = u_k + b_k$; (**B**) ANN supervised learning model implementation. The ANN learns to map a set of signal inputs to specified outputs in the training data. The adaptation (value changes) of the weights is achieved through a cost function for error minimization, and through the training algorithm during the training epochs.

neuron can adopt either a positive or a negative value. Second, the weighted input signals are summed. Third, an activation function $\varphi(\bullet)$ limits the amplitude range of the output of the neuron to a finite value. For example, the typical output of a neuron falls within either the closed interval [0, 1] or alternatively [–1, 1]. Many types of activation functions can be used to train an ANN.[113,114] The model also includes an externally applied bias, denoted by b_k. The bias b_k has the effect of increasing or lowering the net input (u_k) to the activation function, depending on whether it is positive or negative, respectively.

The architecture (i.e. layout and connections between the neurons) that can be adopted into the design of an ANN depends on the task and the function for which the system is designed. In the most common

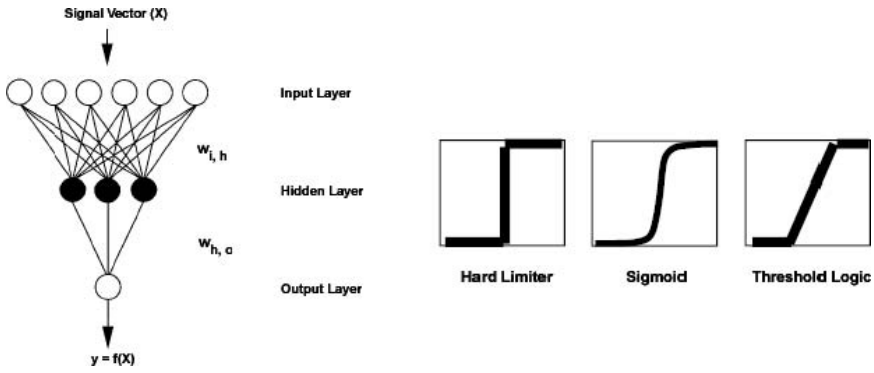


Fig. 2. is a schematic representation of a feed-forward MLP. Computational nodes (neurons) are represented as circles with weighted inputs and output shown as arrows. Also shown, are the three common types of nonlinearity that can be used by the neurons to determine output: hard limiter, sigmoid and threshold.

architecture, the multiple layer perceptron (MLP) (Fig. 2), neurons are classified into three types: 1) *input neurons*, responsible for receiving and passing the signal vector for mathematical processing; 2) *hidden neurons*, in which the information received from the input layer undergoes further processing; and 3) *output neurons*, which produce the system's output or result. The neurons are arranged into layers with corresponding names.

Training the ANN. The network can be trained using either of two types of learning techniques: *supervised learning* or *unsupervised learning*. In the supervised learning paradigm, the system is presented with a set of input vectors that are arbitrarily matched or paired to a set of corresponding output vectors. Once the learning process is over, the trained ANN generates outputs to categorize or approximate, depending on the nature of the task, any given new input vectors. In this type of learning, the ANN is trained with a "teacher." Figure 1(B) shows the basic scheme used for this type of learning. In the unsupervised learning paradigm (training "without a teacher"), the ANN is presented with only input vectors. The ANN then becomes tuned to the statistical regularities in the input data and

gains the ability to form internal representations for encoding features of the input, thus creating new classes in an autonomous fashion. The ability to create internal representations or an internal mapping from input data under unsupervised learning is also known as "self-organized" learning, which is the principal learning paradigm for the training and operation of another form of ANN known as a self-organizing map.

The learning tasks for which an ANN is typically trained include: 1) pattern recognition, whereby the input or received pattern/signal is assigned to one of a predefined number of classes; 2) function approximation, which implies designing an ANN that approximates an unknown function $[y = f(x)]$ describing the input-output mapping; and 3) classification, which typically consists of assigning input vectors to one of two or three classes (e.g. yes or no, numerically mapped to 0 and 1, respectively; or high, medium, and low, numerically mapped to 1, 0, and −1, respectively).

ANN requirements. *Neural Networks and Genome Informatics*, by Cathy H. Wu and Jerry W. McLarty,[115] provides detailed descriptions of the types of learning tasks and biomedical applications of ANNs and the transformation schemes of genomic and proteomic data for vector representation. Their book also addresses some of the requirements and factors that affect the application of ANNs to molecular sequence analysis, such as feature representation, data encoding, analysis of data containing sequences of varying length, and analysis of data sets with limited or missing data. Encoding schemes for protein feature representation, a process usually referred to as *data pre-processing*, and data *post-processing*, are described below.

Data pre-processing. Data pre-processing consists of data encoding and feature representation. Neural networks, like many other machine-learning algorithms, require numerical values for processing. Data encoding addresses the problem of protein representation for mathematical and computational processing. A protein sequence is composed of a series of amino acids represented by the characters A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, and Y. A sequence of alphabetic characters cannot, however, be used in a mathematical computation because none of the elements has a numeric value. Thus, representation of a protein sequence in a numerical and

quantitative way requires generation of a code to represent the corresponding amino acid types. Data encoding can be very intricate and greatly impact the performance of the ANN. A common practice is to encode each one of the 20 letters corresponding to the 20 amino acid types of a protein into a numerical binary scheme. For instance, each letter can be represented by a 20-dimensional binary vector, i.e. 20-bin representation. In such case, amino acids are represented by a set of 19 zeros and a single one uniquely positioned to represent each amino acid. For example, A = 00000000000000000001, C = 00000000000000000010, D = 00000000000000000100, etc. Alternatively, a lower dimensional vector can be generated based on the known physicochemical properties of each amino acid type. Both schemes have their advantages and disadvantages. One of the major advantages of binary representation is that very small changes in amino acid composition between sequences can be easily detected and mapped by the ANN. In fact, this type of encoding representation has been successfully used for other tasks, such as clustering aligned protein sequences.[116] On the other hand, depending on sequence length and amount of samples in the data, 20-bin representation can greatly impact the size complexity of the ANN model. The use of large input layers increases the probability of over training, i.e. the probability that the ANN learns or memorizes insignificant patterns of the training data.[111,117] Meanwhile, encoding schemes based on physicochemical properties usually require lower dimensional vectors for amino acid type representations.[115,118,119] In addition, such schemes allow the encoding of a variety of features, such as hydrophobicity, volume, bulkiness, etc. Furthermore, the number of features that can be encoded does not necessarily impact the size of the vector. Schneider and Wrede (1998) devised an encoding scheme using eigenvalues for each amino acid type from principal components derived by principal component analysis (PCA) of 143 physicochemical scales.[120] Physicochemical values or any combination of features, such as those shown in Table 1, can be assigned to any amino acid type within a sequence. Through either binary or physicochemical encoding procedures, protein sequences are translated to numerical vectors for mathematical processing by the neural network.

Table 1. Schemes for Encoding Amino Acids. (A) binary representations (20-, 6-, and 9-bin), used by Brusic *et al.* (1995), where rep "6" assigns a 6-place string where each place is a scalar value for a feature (hydrophobicity, volume, charge, aromatic side chain, hydrogen bonds) or a correction bit ("a" stands for "10"); "Rep 9" is an intermediate representation using a feature-based grouping of amino acids. (B), used by Milik *et al.* (1998), where every amino acid is represented by six properties (volume, bulkiness, flexibility, polarity, aromaticity, and charge)

Amino Acid	Representation (A)			Representation (B)					
X	"20"	"6"	"9"	Volume	Bulkiness	Flexibility	Polarity	Aromaticity	Charge
A	10000000000000000000	53500a	100001000	0.1677	0.4433	0.2490	0.3951	0.0	0.5
C	01000000000000000000	45500a	000001001	0.3114	0.5506	0.2048	0.7441	0.0	0.5
D	00100000000000000000	a5001a	001001000	0.3054	0.4532	0.8675	1.0000	0.0	0.5
E	00010000000000000000	a60010	001000100	0.4970	0.5567	0.8112	0.9136	0.0	0.5
F	00001000000000000000	185a0a	100100101	0.7725	0.8976	0.0763	0.0370	0.6667	0.5
G	00000100000000000000	525000	000001000	0.0	0.0	1.0	0.5062	0.0	0.5
H	00000010000000000000	565010	010100111	0.5569	0.5632	0.1124	0.6790	0.5556	0.5
I	00000001000000000000	37500a	100010011	0.6467	0.9852	0.6707	0.0370	0.0	0.5
K	00000000100000000000	a7a02a	010000111	0.6946	0.6738	0.6867	0.7901	0.0	1.0
L	00000000010000000000	375000	100010001	0.6467	0.9852	0.2811	0.0	0.0	0.5
M	00000000001000000000	375005	100000100	0.6108	0.7033	0.0	0.0988	0.0	0.5

(*Continued*)

Table 1. (*Continued*)

Amino Acid	Representation (A)			Representation (B)					
X	**"20"**	**"6"**	**"9"**	**Volume**	**Bulkiness**	**Flexibility**	**Polarity**	**Aromaticity**	**Charge**
N	00000000000100000000	655020	000001000	0.3174	0.5156	0.6747	0.8272	0.0	0.5
P	00000000000010000000	555000	100000010	0.1766	0.7679	0.8594	0.3827	0.0	0.5
Q	00000000000001000000	66502a	000000100	0.4910	0.6048	0.7952	0.6914	0.0	0.5
R	00000000000000100000	a8a046	010000110	0.7246	0.5955	0.9398	0.6914	0.0	1.0
S	00000000000000010000	645010	000001011	0.1737	0.3222	0.8514	0.5309	0.0	0.5
T	00000000000000001000	55501a	000001010	0.3473	0.6771	0.5984	0.4568	0.0	0.5
V	00000000000000000100	365007	100011010	0.4850	0.9945	0.3655	0.1235	0.0	0.5
W	00000000000000000010	0a5a10	100100110	1.0	1.0	0.0402	0.0617	1.0	0.5
Y	00000000000000000001	285a18	000100111	0.7964	0.8008	0.5020	0.1605	0.6667	0.5
X	11111111111111111111	555555	111111111						

Feature representation is perhaps the most important aspect of data pre-processing. When analyzing proteins, each letter in the molecular string represents an amino acid. Each amino acid has unique chemical properties associated with specific states or values. A string of characters can be given biological meaning by substituting the characters with their respective chemical values. Many hydrophobic,[121–125] chemical,[126] physical,[127–130] statistical preference,[131] biological[132] and mathematical[133] scales of amino acid attributes have been used. Because of the large number of possible feature representations, data pre-processing of biological sequences can be difficult and confusing. The guide at this crucial step in the design and application of an ANN system should be the problem definition, i.e. determination of the desired input and output mapping for the specified task or goal. The selection of appropriate feature representation and encoding method for a given task limits and specifies the information presented to the ANN. Furthermore, it establishes the parameters of the structures and functions that can improve the accuracy of ANN and allows extraction of statistical consistencies or hidden features in the given sequences. Also, as with the application of any type of statistical analysis or machine-learning technique, the numerical vectors that result from the feature representation and encoding method of the molecular sequences for ANN analysis need to have a relative degree of logic to conform to the basic premise that vectors of similar sequences be close together, and vice versa. This is important if vectors are to carry the biological information of the sequence they represent and maintain both the biological uniqueness and diversity that result from the amino acid composition or sequence length. In contrast, poor feature representation and inadequate encoding methods can result in inadequate vectors, preventing maximal extraction of the statistical features that connect sequences with their structures and functions.

Many sources and tools for performing the numeric transformation of biological sequences are available. In addition to published data, an increasing number of web-based tools can assist researchers in performing these numerical transformations automatically. The Expert Protein Analysis System (ExPASy) from the Swiss Institute of Bioinformatics (http://us.expasy.org/) has several sequence analysis tools and software

packages, including ProtScale,[134] which assigns a numerical value to each type of amino acid as predefined by any of the >50 scales of physicochemical properties and conformational parameters. Thus, substituting the string of amino acid characters with the corresponding listed values or selecting the window size and running the program to compute the profile can transform any given peptide or protein sequence into a numerical/signal vector. Software tools for the automatic modeling of a protein's three-dimensional structure are also available on the Internet. These web-based computational resources are designed to provide atomic coordinates that can be used as a Cartesian three-dimensional representation of a given sequence, which in turn can be used for ANN processing. Examples of web-based servers for homology modeling of three-dimensional structures of proteins include: Geno 3D (http://pbil.ibcp.fr/htm/index.php), from the Institute of Biology and Chemistry of Proteins, Lyon, France; and SCRATCH (http://www.igb.uci.edu/servers/psss.html), from the Institute for Genomics and Bioinformatics, University of California, Irvine.

Data post-processing and computational requirements. Post-processing is the conversion of the ANN output to a different or usable form. It comprises output encoding, which is much more straightforward than pre-processing data encoding. For example, if the task is to classify a signal vector as yes or no, then only one output neuron, mapped to 1 for "yes" and 0 for "no," may be used.

The computational requirements for ANN applications depend on factors such as the type of ANN model, magnitude of the data set, and the learning task. ANN technology has advanced quickly, generating a market of both proprietary and freeware systems that make ANN fairly accessible and easy to implement. In most cases, ANN applications can be carried out on a desktop computer. Shareware and freeware ANN simulator software options include Neural Network Toolbox for Matlab and the freely available Stuttgart Neural Network Simulator (SNNS) (http://www-ra.informatik.uni-tuebingen.de/SNNS/). Additional software can be found at the website http://www.geocities.com/fastiland/NNwww.html, hosted by Professor Andrés Pérez-Uribe, University of Applied Sciences of Western Switzerland, Yverdon-les-Bains; and at http://www-dsi.ing.unifi.it/neural/w3-sites.html#software from

the Machine Learning Neural Networks Group, Department of Systems and Computer Science, University of Florence, Italy.

Feature extraction with ANN. The ANN has been regarded as a "black box" because its operation is neither fully explained nor fully understood. The analysis of a fully trained ANN is in itself considered a science. Nonetheless, compared to traditional statistical approaches, the ANN has considerable strengths and advantages, including the ability to: 1) model complex multidimensional and non-linear relationships, which makes the ANN technique ideal for processing biological information; 2) generalize and find relationships within the data while tolerating erroneous or noisy data; and 3) be retrained with expanded (or new) data sets, which is a plus for model refinement requirements. In fact, an ANN with hidden nodes can extract from input information the higher order features that are ignored by linear models. During supervised training, the ANN is taught to map a set of input patterns to a corresponding set of output patterns. Hidden nodes form internal representation, which is reflected on the weight values of the connections. Post-training analysis of the network weights provide insight into the sensitivity and relevance of input features or attributes for feature selection. New and powerful analytical techniques for weight analysis, also known as contribution analysis, have been developed.[135–138] Although interpretation of the weights of the connections between artificial neurons requires careful consideration, it is generally accepted that the greater the weight value in a connection, the greater the importance of the parameter(s) linked to or associated with that connection. This relationship is analogous to that of the biological counterpart of the artificial neuron, where the strength and number of synapses between biological neurons are believed to be relevant to establishing a particular associative path.[110] In addition, graphical representations of the weights are very helpful and are often the most common form of weight analysis. For example, Holbrook *et al.* (1990) applied ANN to extract information about the surface accessibility of protein residues from a protein sequence using a database of high-resolution protein structures.[139] The ANN was trained to predict the water accessibility of a central residue in the context of its flanking sequence. The protein sequences

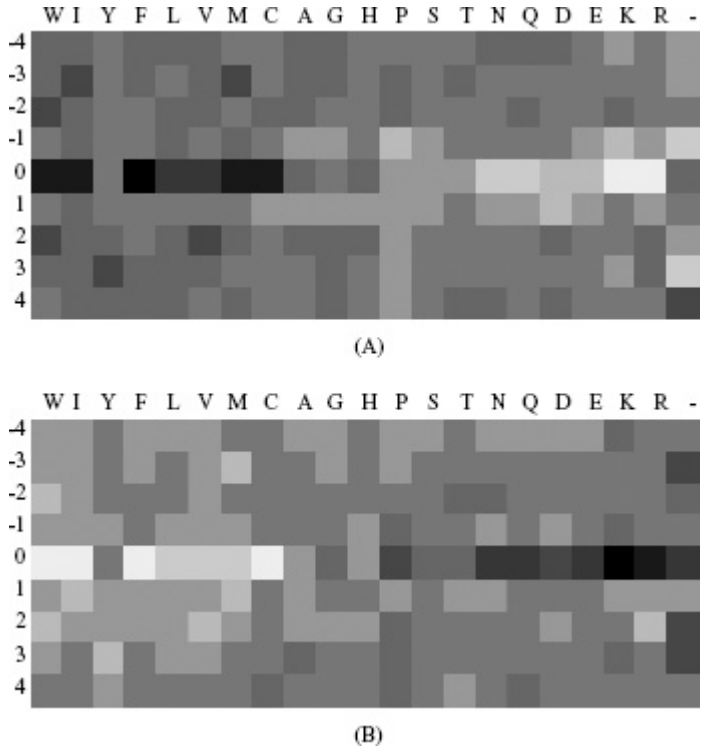


Fig. 3. Weight plots for ANN binary model of surface exposure (from Holbrook (1990)).[139] (**A**) is the weight matrix for the buried residue predictions; (**B**) is the matrix for the exposed residue predictions. For each window position and amino acid type the weight of its connection (link) to the next layer is represented as a shade of gray. Darker shades indicate higher weight values. The amino acid residues are in order of decreasing hydrophobicity according to Eisenberg scale.[140]

were presented to the ANN as sliding windows of variable sizes (up to 13 residues) centered on the amino acid of interest. The analysis was performed by scanning the entire sequence with these sliding windows. To predict the surface exposure of the protein residues, the investigators defined two categories for the buried (inaccessible) and exposed (accessible) residues. Figure 3 shows a graphical representation of the weights in the ANN after training. Examination of the

ANN weights allowed identification of the major factors influencing residue exposure. Such analysis made apparent that hydrophobic residues two or three amino acids up or down from the central amino acid residue hinder its surface exposure. In general, they found that high weights for neighboring hydrophobic residues correlated with their trend to favor burial of the central residue, while high weights for neighboring hydrophilic residues correlated with their trend to favor exposure of the central residue.

Design of Biomolecules with Desired Antigenic and Immunogenic Properties with Artificial Neural Networks

ANNs have been successfully applied to a wide range of disciplines, including aerospace,[141,142] robotics,[143,144] medicine,[145,146] and financial[147] and weather[148] forecasting. This technique, however, has only recently been used in molecular biology for quantitative structure-activity relationships (QSAR) determination.[149–154] A major focus of research aimed at predicting protein QSAR from sequence alone is based purely on empirical approaches. That is, most of the research is based on the use of databases of proteins with known structures and functions. The aim is to find relevant features in these databases, which can then be extracted and used to derive rules. The ANN can use these rules to generalize QSAR models for proteins of unknown functions. Such an approach remains one of the most practical starting points to gain the knowledge needed for protein design. The following sections describe some examples of how researchers have applied the ANN technique for the design of molecules with intended immunological or antigenic properties, particularly, the design of T-cell epitopes. Also, computer-based strategies for peptide/protein and "artificial" protein design will be discussed.

T-cell Epitope Prediction

T cells of the immune system continually check for the presence of foreign antigens that may indicate the presence of invading microorganisms

(e.g. viruses, bacteria, fungi, parasites), mutated self-cells (e.g. tumors), or non-self cells and tissues (e.g. transplants). These foreign and self-proteins are processed through specialized mechanisms and degraded to short peptides that are presented on the host cell surface for recognition by the T-cell receptor (TCR).[155,156] The process involves major histocompatibility complex (MHC) molecules, which bind short peptides and display them to T cells. Peptides presented by MHC molecules originate from intracellular (MHC class I) or extracellular (MHC class II) proteins.[157] MHC class I bound peptides activate cytotoxic T cells, resulting in killing of target cells (e.g. infected, neoplastic, or transplanted tissues), while MHC class II bound peptides serve mainly in regulation (initiation, enhancement, and suppression) of immune responses.[157] The ability of the immune system to respond to a particular antigen varies based on an individual's pattern of MHC genes. In humans, MHC molecules are known as human leukocyte antigen (HLA) class II and I. Each person has only three to six HLA class I molecules and at least that many HLA class II molecules.[157] MHC genes show extensive polymorphism; >800 variants of HLA class I and >500 variants of HLA class II molecules have been characterized.[158,159]

MHC molecules have a peptide-binding groove that can bind peptides in a seemingly non-selective and unrestrained manner.[160] For example, the number of peptides that can bind to each HLA class I molecule has been estimated at 1,000–10,000 sequences, or an average of 0.1%–5.0% of all overlapping 9- and 10-mer peptides spanning the entire protein sequence.[161] Binding of a peptide to an MHC molecule is a prerequisite for recognition by T cells, but only certain peptides can bind to a given MHC molecule.[162] However, binding in itself is not enough. Some peptides are good MHC binders in biochemical assays, but do not elicit a T-cell immune response.[161,163] probably because they are not properly processed and presented by MHC molecules. Thus, T-cell epitopes are a subset of MHC-binding peptides.[163] Determining which peptides bind to a specific MHC molecule is crucial to understanding the mechanisms of immunity and for vaccine and drug design.[164] Fortunately, binding affinities of thousands of MHC ligands have been tested and characterized. Results are available in

databases such as MHCPEP,[165] JenPep,[166] IMGT/HLA,[158,159,167] SYFPEITHI[168] and MHCBN.[169] Such a vast amount of information serves as the basis for empirical computational approaches to elucidate the physicochemical and structural parameters that regulate peptide binding to MHC molecules.

History of T-cell epitope prediction. Algorithms for prediction of T-cell epitopes can be classified as direct or indirect methods.[170] Direct methods are based on information about the T-cell epitopes, whereas indirect methods are based on information about the MHC-binding peptides. In the past, direct prediction methods relied on the identification of amphipathic structures.[171–174] The AMPHI algorithm is based on this assumption.[175] Another algorithm, SOHHA, was developed based on the assumption that T-cell epitopes consist of three to five helical turns with a narrow strip of hydrophobic residues on one side.[176] Algorithms based on secondary structure analysis were superseded by the identification of common motifs among T-cell epitopes in primary structures. Rothbard and Taylor (1988) collected about 57 T-cell epitopes and, based on their patterns, published a list of motifs.[177] They developed an algorithm based on the association between cysteine-containing T-cell epitopes and certain other residues. The system searches for triplets, including CAK, CLV, CKL and CGS, in the peptide sequence.[178] In 1995, two computational T-cell epitope prediction tools, EpiMer and OptiMer, were developed based on knowledge of MHC binding motifs.[179] OptiMer predicts amphipathic segments of protein with high motif density, and EpiMer locates segments of protein with high motif density. However, when tested against databases of human and murine T-cell epitopes, these direct prediction methods performed with low accuracy.[180] In the last decade, several indirect methods have been developed that predict MHC binders instead of T-cell epitopes.[181–184] Due to more specific interactions of MHC molecules and peptides, these methods are more accurate than direct T-cell epitope prediction methods.[170] These indirect methods are based on: 1) structural comparative[185] and *Ab initio*[186] modeling; 2) binding motifs[162,181–184,187–190]; 3) quantitative matrices (QM)[191–199]; and more recently, 4) machine learning techniques like neural networks.[163,170,200–202] The application of ANN algorithms has

been reported to outperform binding motif and QM-based methods.[161,170,203] Predictive performances using ANN methods for the prediction of MHC-binding peptides are in the order of 80% in sensitivity and specificity.[161,163,204]

Prediction of MHC class I and class II binders with ANN. Brusic and colleagues (1995) were the first to report on ANN methods for prediction of MHC-binding peptides.[119] They trained a fully connected three-layer back-propagation ANN to classify peptides into those predicted to bind and not to bind to either human HLA-A2 or mouse H-2K^b. The number of input nodes was equal to the length of the input signal (nine nodes for 9-mers or eight nodes for 8-mers representing HLA-A2 and H-2K^b, respectively); the number of hidden neurons was nine, and one output node predicted binding versus non-binding. Training samples were obtained from the MHCPEP database. The training set consisted of 186 positive examples (binding peptides) and 1,071 negative examples (non-binding peptides) for the HLA-A2 predictions; and 30 binders and 796 non-binders for the H-2K^b predictions. For the encoding of peptide vectors, three signal input representations (Table 1) were used: a 20-bin representation in which each amino acid was encoded by a unique binary string of 0 and 1; Rep6, a six-number representation based on a scalar value for a physicochemical feature; and Rep9, a nine-number intermediate representation using a feature-based grouping of amino acids. During ANN training, each signal input in the training set was mapped to corresponding binding affinities, represented by IC_{50} values, which were defined as the concentration of peptide required to inhibit the binding of a universal DR4-binding peptide by 50%. The system was validated by testing the ANN performance with selected 9-mers from published data obtained by *in vitro* assays.[195] The evaluation set consisted of 60 known binders and 68 non-binders for HLA-A2 and 23 binders and 77 non-binders for H-2K^b. Of the 128 peptides in the evaluation set for HLA-A2 predictions, 30 were derived from natural proteins and the remainder from synthetic peptides, including polyglycine and poly-alanine sequences. The H-2K^b predictions used two test groups: one with nine known binders from different proteins, and the other with 14 binders and 77 non-binders from natural proteins.

The ANN had a predictive accuracy of 78% for binding to HLA-A2 and 88% for binding to mouse H-2K^b.

Milik and colleagues (1998) applied the ANN method to the prediction of mouse class I H-2K^b using 8-mer peptides from a phage display library.[205–207] Binding properties were measured by a competitive CTL assay and values expressed as the amount of competitor peptide (OVA 257) necessary to induce 50% inhibition.[208] They obtained sequences of 181 binding and 129 non-binding phage. For the classification procedure, the data were divided into three sub-sets: training set consisting of 93 binding and 130 non-binding peptides; a validation set 1 (18 binding and 13 non-binding), and validation set 2 (33 binding and 23 non-binding). The validation set 1 was used for fine-tuning the system during the learning phase, and validation set 2 was used for evaluating the performance of the three-layer feed-forward back-propagation ANN. They tested two ANN models based on two different representation schemes for encoding amino acids. The first scheme (ANN1) consisted of a 10-number representation based on the hierarchical organization of amino acids[209]; the second (ANN2) consisted of a six-number representation based on normalized physicochemical properties of amino acids using volume, bulkiness, flexibility, polarity, aromaticity, and charge. ANN1 consisted of one input layer with 180 nodes, one hidden layer with five hidden nodes, and one single output neuron, whereas, ANN2 consisted of 48 input neurons, 3 hidden neurons, and 1 output neuron. To evaluate the validity of the analysis and trained ANN, they predicted H-2K^b binding peptides from the sequence of chicken albumin, a protein with a well-characterized H-2K^b epitope. All peptides that were classified as "binding" were synthesized and tested for experimentally. ANN1 and ANN2 selected the majority of binding peptides tested, with sensitivities of 70% and 80%, respectively.

Honeyman *et al.* (1998) and Brusic *et al.* (1998b; 121–130) reported high accuracy in the prediction of binding peptides (9-mers) to the class II HLA-DR4(B1*0401). Their approach was different from those previously discussed for the class I alleles. In both cases, a hybrid method was developed for predicting binding peptides to MHC class II molecules. Their method, termed PERUN, was implemented

by combining available experimental data and knowledge of binding motifs, quantitative matrices (QM) for pre-processing, an evolutionary algorithm (EA) to derive alignment matrices, and an ANN to perform classification. For development of this hybrid system, peptide sequences and their binding affinities were collected in the data extraction stage from the MHCPEP database, experimental binding data[210] and other private databases (V. Brusic; L.C. Harrison and M.C. Honeyman, unpublished). The initial data set consisted of 650 peptides: 338 binders and 312 non-binders to HLA-DR4(B1*0401). In the pre-processing stage, an EA was used to search for predictive peptide alignment matrices, i.e. quantitative matrices. Through a genetic mechanism (mutation, cross-over, and reproduction) and competition, a population of aligned matrices was generated and selected for improved fitness (characterized by good discrimination between binding and non-binding peptides to HLA-DR4). The fitness function incorporated in the EA algorithm was based on the knowledge of primary anchor positions in reported binding motifs,[189] which was used to fix position 1, corresponding to the primary anchor in each matrix, while the rest of the matrix was subjected to the application of the abovementioned genetic mechanisms. The end product of the EA search was an alignment matrix, where each residue at each position in the 9-mer was assigned a binding score. The aligned matrix was then used to find and align putative 9-mer cores of known binders for the final stage of classification by the ANN. This set of aligned peptides served as the training set, which comprised of aligned binding 9-mer cores (338 peptides) and non-binding 9-mers (312 peptides). The aligned peptide sequences were then transformed to numerical signal vectors using a 20-bin amino acid representation. The ANN system was designed as a three-layer, fully connected, feed-forward back-propagation ANN. Because the binding affinities of the peptides in the training set were not all derived by the same experimental method, affinity measures were grouped into no affinity, low affinity, moderate affinity and high affinity and output values represented as zero, six, eight and 10, corresponding to these groups, respectively. Honeyman *et al.* (1998) reported that the ANN-based binding prediction on an experimental binding affinity of synthetic

peptides was 77% sensitivity and 80% specificity. Meanwhile, Brusic *et al.* (1998b, 121–130) compared the PERUN method with the QM-based and binding motif-based methods for prediction of low, moderate and high affinity binders. PERUN slightly outperformed the QM method, with prediction performance of 73% for low binders, 86% for moderate binders and 88% for high affinity binders, whereas the QM performance was 73%, 82% and 87%, respectively. The binding motif method had the lowest performance (63%, 69% and 74%, respectively). Such an approach integrates the strengths of the different methods and minimizes their disadvantages. For instance, binding motifs encode the most important rules of peptide-MHC binding interaction.[184,189,196,211] However, for class II molecules, binding motifs are less well defined.[189] MHC class II molecules show degenerate motifs, which makes peptide alignment more difficult. Because of this, except for certain molecules,[196,211] binding motif-based methods show poor generalizations. In contrast, QM-based methods, which are in essence refined binding motifs, can predict large subsets of binding peptides reasonably well and can be used when data sets are limited.[192,193,195,196,212,213] However, QM cannot deal with nonlinearity in data and may miss distinct subsets of binding preferences, such as medium binders.[200] Also, QM methods are not adaptive and self-learning, so integration of new data usually requires redesigning of the alignment matrix. On the other hand, ANNs can deal with nonlinearity and are adaptive and self-learning, but usually require a large amount of data.

A limitation of ANN predictions based on the indirect methods described above is their inability to discriminate between T-cell epitopes and non-epitope MHC binders, mainly because these methods only predict the MHC binders from antigenic sequences. Bhasin and Raghava (2004) recently developed a direct method for predicting cytotoxic T lymphocyte (CTL) epitopes from antigenic sequences using a hybrid method that combines QM, support vector machine (SVM), and ANN methods. Their system is based on 1,137 experimentally proven CTL epitopes and 1,134 non-epitopes (786 non-epitopes 9-mers and 348 MHC non-binders). The system was evaluated with a blind data set consisting of 63 CTL epitopes, 63 non-epitope

MHC binders and 63 MHC non-binders. ANN and SVM outperformed the QM method. For training of both the ANN and SVM, peptide sequences were changed to numerical vectors by encoding sequences with 20-bin representation. The ANN performance was 72.2% accuracy, 73.2% sensitivity, and 71.2% specificity. The SVM performance was 75.4%, 73.8%, and 77.0%, respectively, while the QM method performance was 70.0%, 65.2%, and 74.9%, respectively. Also, the ANN and SVM were used to perform consensus and combined prediction of CTL epitopes. In consensus prediction, epitopes predicted by both methods were considered "epitopes," otherwise they were considered as "non-epitopes." In combined prediction, epitopes predicted by either method were considered as "epitopes." In addition, two models of such implementation can be used, i.e. in one model the ANN can be used as the base method, while in the second model the SVM is used as the base method. The CTLPred server (available at http://www.imtech.res.in/raghava/ctlpred/) is based on such an approach. They also developed the ComPred method[169] based on a hybrid method using ANN and QM filtering for the combined prediction of MHC binding peptides or CTL epitopes. Using this combined prediction approach, they achieved a sensitivity of 79.4% and a specificity of 88.4% for discriminating between T-cell epitopes and non-epitope MHC binders. Another of their programs, ANNPred, uses a feed-forward back-propagation ANN for the prediction of MHC binders to 30 MHC alleles. Both ComPred and ANNPred are available at http://bioinformatics.uams.edu/mirror/nhlapred/index.html.

Data encoding for T-cell epitope and MHC binding predictions with ANN. Representation of peptide sequences using 20-bin numerical representation (0s and 1s), appears to provide an ANN with sufficient discriminative power to predict (or classify) CTL epitopes[169,170] and MHC binding activities of peptides.[119,163,214] However, this sparse encoding of the amino acids ignores their physicochemical similarities or dissimilarities. Furthermore, in some cases, selection of numerical representation schemes for signal processing is reported to have statistically significant effects on ANN performance.[201,202,215] As previously discussed, Milik *et al.* (1998) reported that an ANN derived

from the individual representation of amino acids using physicochemical properties (ANN2) outperformed an ANN derived from the clustering of amino acids (ANN1). Also, Nielsen *et al.* (2003) reported that an ANN trained with a novel sequence encoding scheme using Blosum50 and hidden Markov model sequence encoding for prediction of T-cell class I epitopes was superior to one trained on 20-bin sequence encoding (accuracy of 91.2% and 87.7%, respectively). Another strategy to enhance prediction performance is to combine several ANNs derived from different sequence-encoding schemes instead of using one ANN system derived from a single sequence-encoding scheme.[202,215] Although in some cases using either 20-bin representation or physicochemical features may not affect classification performance, the latter encoding strategy has important advantages over sparse encoding: 1) it allows researchers to weigh in the contribution of specific amino acid parameters and directly extract their correlation to MHC peptide-binding properties and CTL epitope SAR; 2) continuous-property scales for signal encoding, unlike sparse encoding, reduces the dimension of the vector input and the danger of generating arbitrary correlations that may lead the ANN to overfit (this is especially important when data are sparse); and 3) in contrast to sparse representation, physicochemical or encoding schemes, like Blosum50, provide the neural network with less precise and more general information about the sequence, which allows the system to generalize better, mainly because it encodes subtle chemical and evolutionary relationships between the 20 amino acids.[202]

Relevance of T-cell epitope prediction by ANN methods. The experiments described above provide compelling evidence for two possible applications of ANN technology for immunological therapies involving cytotoxic T-cell and T-cell mediated immunoresponses: vaccine and drug design. These applications differ, however, in their requirements for specificity and sensitivity. For vaccine design, it is more important to predict all possible binders to a given MHC molecule (specificity), whereas for drug design it is desirable to predict those peptides that will bind with highest affinity (sensitivity). It should be possible to manipulate the performance of the ANN to optimize specificity or sensitivity as required for a given application.

Towards this aim, in addition to the programs already mentioned, other systems are available to researchers. The MULTIPRED program[216] (available at http://antigen.i2r.a-star.edu.sg/multipred/) is based on ANN methods to predict HCV T-cell epitopes for MHC class I A3 molecules and human papillomavirus (HPV) type 16 T-cell epitopes for MHC class A2 variants, and is reported to achieve good sensitivity and specificity and prediction capability, with an area under the receiver operating system curve (A_{ROC}) >0.80. NetChop server[215] (available at http://www.cbs.dtu.dk/services/NetChop/) uses ANN method to predict proteasome cleavage sites responsible for generating CTL epitopes

ANN-based predictions can also reduce the number of bench experiments needed for T-cell epitope screening. ANN applications allow investigators to reduce and focus the number of candidates for identification of T-cell epitopes while increasing the efficiency of the search space for these candidates. For example, Honeyman and collaborators (1998) synthesized 68 peptides (9-mers) and used a trained ANN to perform binding prediction on this set. After completing experimental binding and T-cell response tests, they show that ANN-predicted binding would have reduced the number of synthesized peptides from 68 to 26, with the potential loss of four epitopes. In another case, Schönbach and collaborators (2002) applied ANN methods to the large-scale identification of HIV T-cell epitopes from Pol, Gag, and Env sequences for HLA-A*0201 and found 890 HIV-1 and 232 HIV-2 epitope candidates. The overall sequence coverage of the predicted A*0201 T-cell epitope candidates was 2.7% for HIV-1 and 3.0% for HIV-2. By comparing the ANN-predicted binding with other bioinformatics methods and experimentally confirmed A*0201-restricted epitopes, they were able to extrapolate their results and estimate that approximately 247 predicted HIV-1 are yet to be discovered as active A*0201-restricted epitopes. Finally, with proper encoding, ANNs can also be used for feature extraction. Conclusions can be drawn regarding the relative importance to binding of specific positions in the peptide and chemical properties of amino acids at those positions. Thus, use of these capabilities can facilitate the *de novo*-driven design of viral vaccine or immunotherapeutic candidates.

Peptide and Protein Design

Empirical methods for building predictive models of the relationships between molecular structure and functions of molecular candidates are becoming increasingly important for vaccine discovery and development. Automation of chemical synthesis, tools for designing molecule libraries and high-throughput technology for biological screening experiments robustly and quickly enable target candidates to be discovered more effectively. As vaccine and diagnostic candidates progress down the development pipeline, the ability to predict their immunological, antigenic, pharmacokinetic, and toxicological properties is becoming increasingly important in reducing the number of expensive late-development failures. As a result, the application of ANN methods in this area of research has gained popularity.[217–226] The major rationale for the use of ANN methods for the prediction of antigenic/immunogenic properties of viral molecules is that determining for which structural parameters the ANN best predicts may lead to the identification of relevant features responsible for the observed activities. Extracting such knowledge from primary structure is necessary for the successful design and engineering of molecule candidates with desired properties. In addition, QSARs are usually complex and nonlinear, and the ANN is often chosen because the system tends to provide more flexibility and adaptability for modeling non-linear relationships and is more noise-tolerant compared to other methods.[217]

Schneider *et al.* (1998(b); 12179–12184) proposed a systematic, focused strategy for peptide design through a combination of rational and evolutionary approaches.[227] They described the use of evolutionary strategies as a cyclic-variation selection of peptide candidates followed by ANN modeling of quantitative structure-activity relationships. They reported successful application of an approach to the *de novo* design of a peptide that fully prevents the positive chronotropic effect of anti-b_1-adrenoreceptor autoantibodies from the serum of patients with idiopathic dilated cardiomyopathy (DCM). Based on previous reports showing that short synthetic peptides encompassing the natural epitopes were able to neutralize the chronotropic effect of the autoantibodies, their idea was to test

whether it was possible to derive novel (artificial) epitope sequences directed against the second extracellular loop of human b_1-adrenoreceptor for potential immunotherapeutic agents. Their design strategy is described in the following four steps.

(1) *De novo design and evolutionary approach.* The researchers first identified a sequence with a desired activity (b_1-adrenoreceptor loop 2 epitope) that would serve as a "seed" sequence around which to develop an *in silico* library. A common approach for such searches is through random screening. To obtain a seed peptide, parts of the sequence of human b_1-adrenoreceptor encompassing loop 2 were analyzed by epitope mapping. The ability of the peptide fragments to bind to human anti-b_1-adrenoreceptor antibodies was measured by ELISA. The amino acid sequence ARRCYNDPKC (positions 107–116) was identified as a natural epitope with specific affinity to the antibodies and was therefore selected as the seed peptide.

(2) *Focused library strategy.* Using the amino acid sequence ARRCYNDPKC as the seed peptide, a set of 90 variants were generated by a simple algorithm describing each residue by the respective property values for hydrophobicity[123] and volume.[228] Thus, a total of ninety 20-dimensional vector representations of variants, in terms of these two properties, were generated. These vectors were generated using the Box–Muller formula, and were approximately gaussian-distributed around the seed peptide vector (standard deviation (σ) = 0.1); where g is a gaussian-distributed random number and i and j are random numbers in the [0, 1] interval:

$$g = \sigma \sqrt{-2 \ln(i)} \sin(2\pi j). \qquad \text{Eq. (1)}$$

The rationale for obtaining vectors both close and distant from a seed vector is that peptides with an activity similar to that of the seed peptide are expected to be in close proximity to the seed peptide in sequence space,[229] while those that are more distant are expected to have decreased activity or provide new seed vectors of higher activity. The strategy is designed, through a mathematical approach, to focus

the search for a desired property on a list of peptide candidates that are more likely to conform to such properties rather than following the conventional random search (epitope mapping) for natural epitopes, which is expensive and time consuming. After generation of the *in silico* library, the property vectors were translated back into amino acid sequences by selecting the most similar residues at each sequence position according to their physicochemical properties. The computer-generated variants were synthesized and tested for their ability to bind to human anti-b_1-adrenoreceptor antibodies. The *in vitro* assays confirmed the applicability of the mathematical strategy to the generation of a small focusing peptide library. Several non-natural peptides were found to have increased antibody-binding ability when compared with natural epitopes. Thus, the generation and selection of candidates from a focused library provided an efficient starting point from which to extract knowledge. This approach also allowed them to expand the search to a large number of sequences more cost effectively (*in silico*) and in less time (computational process) than would have been possible with conventional techniques such as random mapping or DNA shuffling.

(3) *Establishment of QSARs with ANN methods.* Armed with the information generated from the focused library, a three-layered, feed-forward network with a sigmoid hidden-unit and a single linear output unit was, under the supervised paradigm, to extract the underlying QSAR by mapping the sequence vectors (90 peptides and seed peptide) to their respective semi-quantitative absorbance properties (defined as high, medium and low binding activities). The trained network was then used as a fitness function for selection of subsequently evolved peptide candidates.

(4) *Evolutionary/mathematical search for new variants with desired properties.* Once the neural network was trained on the information generated from the focused library data, it was used as a heuristic fitness function and focused approach for searching in sequence space. Generally, the expanded sequence space search for the *de novo* design of peptides starts by querying computer-generated sequences. Schneider *et al.* (1998(b); 12179–12184) opted for an evolutionary

strategy for generating peptide libraries. In a cyclic or iterative process, virtual peptide libraries are generated by variation of a seed sequence (parent), the activity of each peptide (offspring) is predicted by the ANN, and the peptide (child) with the highest predicted activity is selected as the parent for the next cycle. This is repeated until no further optimization can be observed. To evaluate the ANN, the selected peptides were tested for their activity and compared to the predicted activities. Despite some conflicts in the degree of binding activity, all peptides predicted as active by the ANN showed activity when tested. The most active peptides shared only one common residue, aspartic acid, at position 7 with the parent or seed sequence. Furthermore, a peptide predicted by the ANN as being the most inactive peptide, which contained favorable binding motifs in two positions, in fact tested to be non-binding. Most striking was the observation of the difference in activities of the *de novo* peptides when compared to natural-derived peptides at different concentrations. Some of the ANN-designed peptides had similar or significantly higher activities at concentrations 10-fold lower than those of natural peptides.

This example for peptide/protein design shows that a mathematical approach using ANN methods to derive knowledge and search in sequence space can be successful for the identification of mimetopes (antigens mimicking a natural epitope sequence) with significantly different residue sequences. It also shows that the most active peptide represents a good starting point for the expanded search and further optimization. One may describe the virtual strategy presented here as the mathematical "molding" of molecular structural features to a desired function. Thus, computer-based strategies for *de novo* molecule candidate design have emerged as a complimentary approach to high-throughput techniques and should prove beneficial for development of immunotherapeutic and diagnostic candidates.[166,230–232]

Other computer-based techniques have been used for the sequence-oriented design of focused libraries. Wrede *et al.* (1998) developed two computer-based techniques for the generation of artificial *E. coli* signal peptidase I (SP I) cleavage target sites. The first

technique (PROSA, "protein sequence analysis") used a simple statistical procedure based on the frequencies of residue classes in a set of aligned amino acid sequences. It generated consensus sequences with a physicochemical property pattern thought to constitute an underlying functional or structural feature in common. The second design technique (SME, "simulated molecular evolution") used ANNs and an evolutionary algorithm for peptide development. The neural networks were used to extract essential cleavage site features from a set of known cleavage and non-cleavage site examples. After successful feature extraction these systems provided a mathematical model of SP I target sites that was used as a heuristic fitness function for a systematic evolutionary search in sequence space.

Artificial Peptide and Protein Design

An interesting idea that emerges from these strategies for *de novo* peptide/protein design is the possibility of designing truly artificial candidates. Since knowledge of QSAR can be extracted from the neural network, it is possible to include non-natural (artificial) residues in the design strategy. This can lead to peptide/protein-like structures that are suitable as leads for immunotherapeutic and diagnostic development. Certainly, if successful, protein engineering will break loose from the natural architectural framework constrained to a 20 amino acid domain. In fact, *de novo* design of peptides/proteins with desired functions should be approached in the same way that modern engineering has modified and changed our environment.

The first step in designing a building is to experiment with many different ideas for the basic form, volume, and the structure of the building to meet its pre-specified function. Added to this architectural designing process is the design of raw materials to specification. Raw materials like iron and aluminum, or building blocks like cinder blocks and bricks are physically and chemically modified to specific strengths and molded to predetermined shapes. All of these factors are needed to achieve the final structure and function for which a building has been designed. Similarly, amino acids, the raw material and building blocks

of proteins, should be modified to specification to achieve favorable structural features for a predetermined function. The mathematical abstraction of favorable property features for the design of artificial amino acids, along with the testing of these engineered peptides/proteins centered on a particular function(s), could make possible smart and fast engineering of biomolecules with pre-specified function. To search and test for potential artificial antigens/immunogens, the peptide/protein library synthesized with modified amino acids would be mixed with the viral particle of interest. Those that bind (*win*) would stick to the viral particle, and those that do not (*lose*) would be washed away. The winning ligands can be selectively used as seed candidates for another round to build more *in silico* lead libraries. With enough candidate samples, their activity-based information can be used to train the ANN. ANN-selected (predicted) candidates can then be validated experimentally. Thus, the ANN "knowledge" and the peptide/protein library, evolve through the iterative process: predicting binding properties and building *in silico* libraries from ANN-selected "seed" candidates. Such a strategy may allow researchers to develop agents that cover the entire sequence space of the virus. It may also help generate more stable and long-lasting agents because, in contrast to peptides/proteins made from the naturally occurring amino acids, modified amino acids may not be as quickly recycled by the body.

Schultz and colleagues have been able to expand the genetic code by building new building blocks (primarily in *E. coli*) for site-specific incorporation of artificial amino acids into peptides/proteins.[233–239] Presently, more than 30 artificial amino acids have been genetically encoded through unique triplet,[234,235] quadruplet,[235,240–243] and quintuplet[244,245] codons. Artificial peptides have already been successfully tested for their binding affinities to proteins of the MHC class I in T cells. Krebs, Folkers and Rognan (1998) synthesized and modified peptides to optimally bind to the MHC I molecule HLA-B*2705 by filling a hydrophobic binding pocket (pocket D) with artificial aromatic amino acids (alpha-naphthylalanine, betanaphthylalanine, and homophenylalanine), whose side chains in position 3 had been altered.[246]

Other Computational Strategies for Design of Immunogens

Evolutionary Algorithms (EAs)

"Evolutionary algorithm" is a generic term that groups three independently developed computational problem-solving methods: genetic algorithms (GA), evolutionary strategies (ES), and evolutionary programming (EP). These three methods are biologically inspired by the principles of Darwinian evolution, but differ in the implementation of these principles. In general, all three methods include treatment of proposed possible solutions as members of a population that are varied for fitness (or adaptation) to their environment. Members of the population are subjected to selection pressures, and the survivors (parents) breed offspring (children) by the application of genetic operations such as mutation, crossover (recombination), or both. These operations proceed in a cyclical manner for a number of user-defined generations, after which fitter population members have evolved from the original population as seen in Fig. 4.

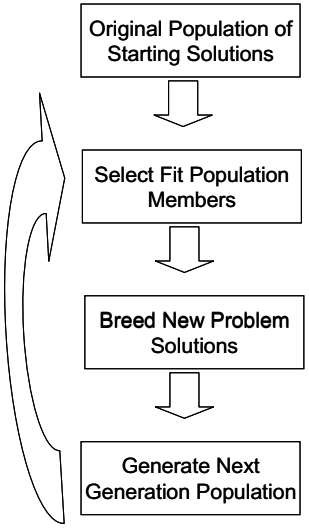

Original Population of Starting Solutions
Select Fit Population Members
Breed New Problem Solutions
Generate Next Generation Population

Fig. 4. General scheme used by EAs for problem solving.

The implementation of any EA begins with the encoding of the problem solutions. Binary encoding is frequently used, but, as discussed earlier, real number encoding (e.g. using physicochemical properties) provides a more natural representation of the problem. Regardless of the encoding type, a coded problem solution is referred to as the *genotype* and can be used to determine the decoded phenotype. The *genotype* can also be called a *chromosome*, and consists of a

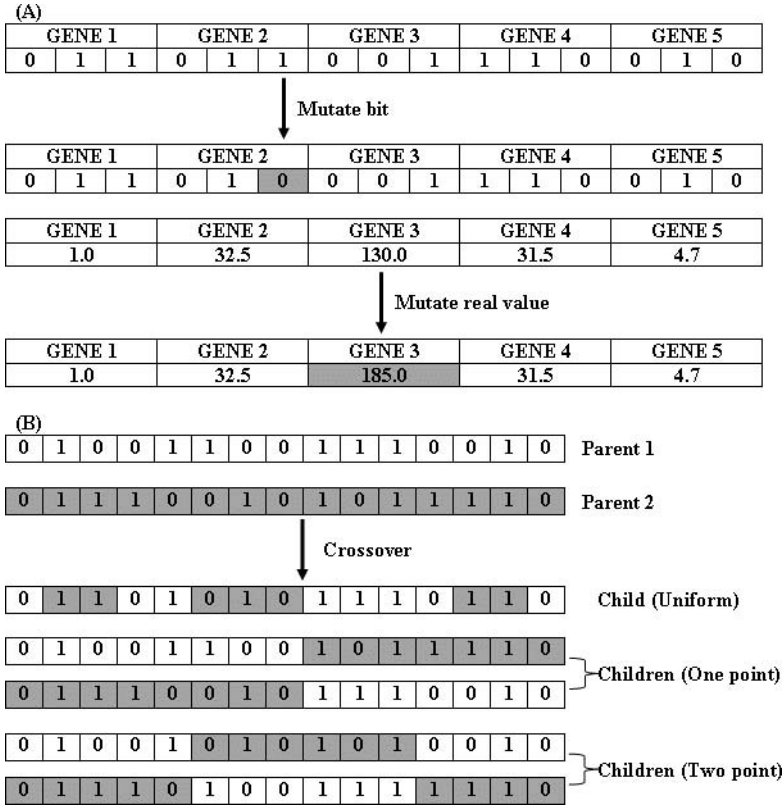


Fig. 5. (**A**) Binary and real-value point mutation operations in a chromosome with five genes representing hypothetical biomolecule conformations; (**B**) breed populations (children) of three different types of crossover operations (uniform, one point and two point). Color shades indicate the parts of the genes coming from respective parents.

collection of *genes* (Fig. 5). The mutation operator is one of the evolutionary methods that can generate offspring solutions for the next generation. Mutation is the only evolutionary operator in use in EP, is the primary operator used in ES, and is a secondary operator in GA applications. The implementation of mutations depends on the representation. Mutation in binary representation presents problems by selecting a bit and randomly changing it to 0 or 1 (independently of the original value of that bit) or by inverting the value of bits selected for mutation. Mutation in real-valued representations requires a random change in the value of the real number encoded at a selected position in the genotype. The crossover operator is the second evolutionary operator used to generate offspring from selected parents. Crossover is the primary operator in EP, a secondary operator in GA, and not used at all in EP. Crossover combines genetic information from two parents to generate one or two children that have features from both parents. Crossover can be implemented in a number of ways (Fig. 6(B)). The selection and the next-generation process include a "survival of the fittest" mechanism. The computational implementation of such a mechanism requires a way to calculate the fitness of each member of the population (fitness function) and an algorithm for selecting members of the population that will pass their genes to the next generation. Fitness values should reflect the relative quality of the problem being optimized to direct the evolutionary search toward a more promising solution. Thus, fitness functions for calculating such values must be written specifically for each problem. As a result, the calculation of fitness values often consumes the majority of the time dedicated to solving problems by EA. However, there is no requirement for fitness values to be determined computationally. GA implementation can use experimentally derived values determined from biological assays that can serve as fitness values for population members.[247] There are many types of evolutionary algorithms used for the selection process, but their discussion is beyond the scope of this chapter.

As previously mentioned, Brusic and colleagues (1998(b), 121–130) developed a hybrid method for the prediction of HLA-DR4 (B1*0401)-restricted peptides. In this case, a GA method was

used to search for predictive alignments as a pre-processing step before ANN analysis. The population in this case consisted of a set of alignment matrices that were transformed into a new population by means of a GA and a selection mechanism for fitness improvement. The constraint applied to direct the evolutionary search, derived from knowledge of primary anchor positions reported in binding motifs, consisted of fixing position 1, corresponding to the primary anchor in each matrix, while the rest of the matrix was subjected to the GA. The size of the population of matrices was arbitrary selected and the selection technique used was elitist in that each parent matrix produced two children, an identical copy of itself and a mutant copy, and passed the offspring with the higher fitness value to the next generation. All matrices of the final generation were used to score peptide alignments by assigning a score to each putative binding core within each binding peptide. In each simulation, the alignment scored as highest by the majority of the final generation matrices was selected and passed to the final stage, i.e. ANN training. For chromosome encoding the researchers used real numbers instead of binary representations. They included mutation and reproduction evolutionary operators, but opted to exclude the crossover operator because it resulted in a population of matrices with high similarities that promoted a linear model of peptide alignment. The measure of the predictive power of an alignment matrix was used to define its fitness. Such a fitness function favored matrices that correctly classified non-binders and resulted in a population of matrices in which individual matrices captured disjoint regions in the solution space. The termination criteria (selection of parent for mutation and reproduction operations) was determined in a separate experiment that measured sensitivity, specificity, and average fitness function and in which each population was evolved up to 10^6 generations. The final generation of matrices was used to score potential alignments, from which the highest scoring alignment of binding peptides was selected. As previously described, this alignment, along with putative non-binders, was used for ANN training.

Other types of EAs, such as genetic programming (GP), have also been used to address the problem described by Brusic (1998(b), 121–130). GP is a subclass of GA that is an automated procedure to find

computer programs that best perform a user-defined task.[248] Through a Darwinian natural selection process, GP is capable of evolving computer programs that solve, or approximate the solution, of a variety of problems. GP focuses on exploiting the regularities, symmetries, homogeneities, similarities, and patterns of problem environments by means of automatically defined functions (ADFs). An ADF is a function (i.e. subroutine, procedure, or module) that is dynamically evolved during a run of GP. ADFs were first conceived and developed by James P. Rice and John R. Koza of the Knowledge Systems Laboratory at Stanford University.[249] Klien and colleagues (2000)[247] applied GP methods to the same data set used by Brusic (1998(b), 121–130). For MHC binding predictions by GP, they used nine amino acid descriptors based on physicochemical properties (hydrophobic, positive, negative, polar, charged, small, tiny, aromatic, and aliphatic). GP instructions set in the virtual machine used a straightforward collection of Boolean functions, which allowed expression of complex associations of amino acid residues, locations, and properties. They implemented a sliding window of length nine to view the peptides and moved the window one residue at a time to look for any possible sequence of nine residues within each peptide. Peptide windows that did not include a known anchor residue in position 1 were not tested. Instead of using mathematical instructions, the researchers used only logical instructions and those that queried the type of amino acid in a specific position or the properties of an amino acid in a specified position. Their system implemented genetic operators for mutation, reproduction, and crossover. However, they reported a lower predictive performance (75% accuracy) compared to the >80% achieved by Brusic (1998(b), 121–130) using a hybrid method. Details of their GP implementation and source code are available at http://www.dd.chalmers.se/~f97johan/Projects/EvComp/index.html.

Probabilistic Graphical Models

Probabilistic graphical techniques result from the merging of probability and graphical theories. A wide variety of probabilistic graphical methods have been developed, including the hidden Markov models (HMM) and Bayesian networks (BN). In fact, systems that use the

mathematical technique known as Bayesian inference have improved the performance of many AI programs to the point that they are being used in the real world. For example, the Microsoft Office 97 paper-clip assistant is based in Bayesian networks (marking Microsoft's first commercial deployment of Bayesian models).[250] The application of BN methods is steadily on the rise in biomedical research. They have been used for protein folding and secondary structure predictions,[251–253] protein-protein interactions,[254–256] gene-drug dependencies,[257] identification of transcription sites[258] and splice sites,[259] estimation of gene regulatory networks,[260–263] and a range of other areas including medical diagnosis.[264;,265]

A Bayesian network (also known as a Bayesian belief network, probabilistic causal network, chain graph, and knowledge map) is a form of artificial intelligence designed to cope with uncertainty.[266] Such systems operate using "Bayesian reasoning" under the Bayes rule principle.[267–270] The BN method has both quantitative and qualitative attributes. With respect to the latter attribute, a BN is represented as a directed acyclic graph in which the nodes represent random variables of interest and the edges (arcs) linking them represent causal influence. In the simplest form of BN, quantitatively each variable is considered to be independent of its ancestors given its parent, where the parent/ancestor relationship is with respect a fixed topological or hierarchical ordering of the nodes. The BN represents a local, rather than a global, probability model, in which each node consists of a set of conditional independent probability distributions (Fig. 6). Among the advantages of Bayesian networks is their ability to: 1) combine highly dissimilar types of data (i.e. continuous or discrete data; or data from disparate sources) into a common probabilistic framework; 2) provide an intuitive and interactive graphical model representation that allows the researcher to read true probabilities from the graphical structure to determine the significance of each feature, which can help explain the relevance of peptide/protein structural features for a given problem application; 3) be used to discover data structure and conditional probabilities directly from raw data (*tabula rasa*) through unsupervised learning; and 4) provide a conceptual framework for the clear distinction between causal and statistical parameters and general concepts extending from this distinction, thus providing a proper tool

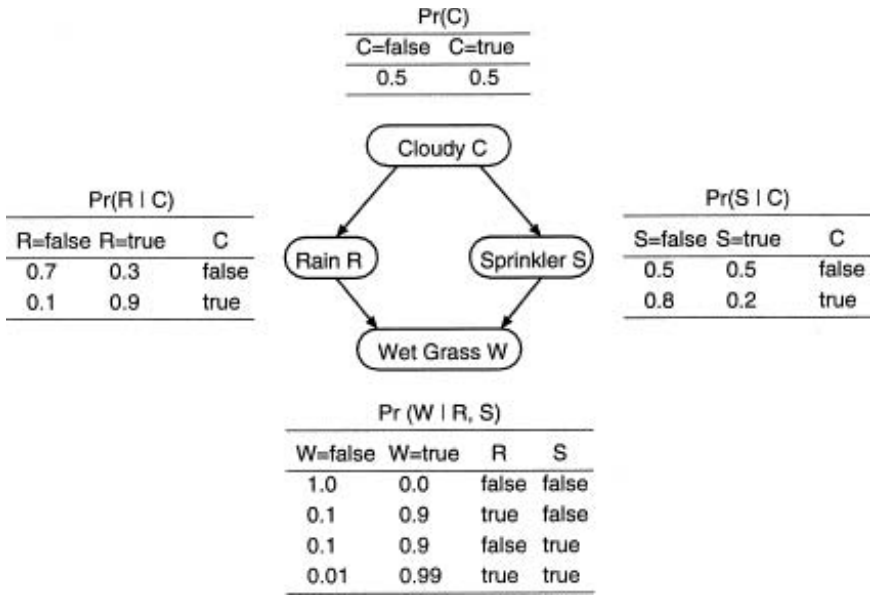


Fig. 6. Example of a well known Bayesian network describing probabilistic dependencies among variables Cloudy (C), Sprinkler (S), Rain (R), and Wet Grass (W). These probabilities form the basis for computing the joint probability: P(C, R, S, W) = P(C) P(R|C) P(S|C) P(W|R, S), that is, the probability of observing a combination of states assuming the specified probabilistic dependencies among the nodes. In this example, the probability of seeing wet grass for a cloudy sky with rain and no sprinkler on is $0.5 \times 0.9 \times 0.8 \times 0.9 = 0.324$. Not all states in the network may be known or observable at any given time. With BNs, marginal probabilities for subsets of known states can be obtained from the joint probability by summing over all possible combinations of states for the unknown variables. For example, if we could not observe *R* or *S*, we would obtain the marginal probability. In this way, the unknown variables *R* and *S* are eliminated from the distribution. Thus, we see that the likelihood of seeing wet grass when it is cloudy is $(0.5 \times 0.1 \times 0.8 \times 0.0) + (0.5 \times 0.1 \times 0.2 \times 0.9) + (0.5 \times 0.9 \times 0.8 \times 0.9) + (0.5 \times 0.9 \times 0.2 \times 0.99) = 0.4221$.

for the investigators to treat non-statistical concepts, such as influence, confounding, effect, intervention, explanation, disturbance, etc.[267]

BN methods has been applied to the prediction of binding peptides to HLA-A2 class I molecules.[271] Astakhov and Cherkasov (2005), based on a data set consisting of 244 HLA-A2-binders (derived from

MHCPEP and SYFPEITHI databases) and a set of 464 non-binders (derived randomly from a human albumin sequence), demonstrated that the BN shows up to a 99.0% accurate identification of the HLA-A2 binding peptides. Peptides were encoded using 20-bin vector representations, and HLA-A2 binding activities representations were the same as those used by Brusic *et al.* (1994(b), 121–130). In addition, their approach may also suggest certain advantages of probabilistic graphical methods over other methods like the ANN. When the training data set was reduced by 40.0% from the original data, the predictive accuracy (as measured by A_{ROC}) of HLA-A2 binding peptides by BN, HMM, and ANN methods was 0.88, 0.85 and 0.85, respectively. Therefore, application of BN methods may provide an alternative approach to immunologists when data are sparse. Astakhov and Cherkasov clearly show that the BN method can potentially serve as a machine-learning tool that can be applied to prediction of viral immunogens as well as to *de novo* design strategies.

The studies described above demonstrate the power of ANN, as well as other machine-learning methods, to exploit information from peptide/protein databases in order to derive SAR and make predictions of peptide/protein functions from primary sequences alone. Most of the studies described herein used three-layered feed-forward ANN with back-propagation methods. Expansion into other ANN methods, such as self-organized maps,[272] self-growing neural networks,[273,274] and associative memory networks,[272] should bring advances in peptidomimetics and *de novo* design. Also, development of hybrid approaches that combine ANN methods with other machine-learning techniques such as HMM, EA, BN, and SVM will enhance prediction performance and advance development of vaccine and diagnostic candidates. Neural network techniques are not just for computer scientists or engineers, but comprise a powerful set of tools for immunologists as well. Advancements in ANN applications to immunology and protein design will depend on the ingenuity in formulating problems and designing neural networks and the reliability of the information stored in databases. We anticipate fast growth in ANN applications for computational virology and in the number of immunologists using such mathematical machine-learning techniques.

References

1. Cunliffe NA, Nakagomi O. (2005) A critical time for rotavirus vaccines: a review. *Expert Rev Vacc* **4**(4): 521–532.
2. Girard MP, Cherian T, Pervikov Y, Kieny MP. (2005) A review of vaccine research and development: human acute respiratory infections. *Vaccine* **23**(50): 5708–5724.
3. Hartaningsih N, Dharma DM, Soeharsono S, Wilcox GE. (2001) The induction of a protective immunity against Jembrana disease in cattle by vaccination with inactivated tissue-derived virus antigens. *Vet Immunol Immunopathol* **78**(2): 163–176.
4. LaBrecque FD, LaBrecque DR, Klinzman D, *et al.* (1998) Recombinant hepatitis A virus antigen: improved production and utility in diagnostic immunoassays. *J Clin Microbiol* **36**(7): 2014–2018.
5. McLean AA. (1986) Development of vaccines against hepatitis A and hepatitis B. *Rev Infect Dis* **8**(4): 591–598.
6. Schwarz TF, Roggendorf M. (1990) Diagnostic of parvovirus B19 infection. *Behring Inst Mitt* **85**: 35–38.
7. Scott RM, Nisalak A, Cheamudon U, *et al.* (1980) Isolation of dengue viruses from peripheral blood leukocytes of patients with hemorrhagic fever. *J Infect Dis* **141**(1): 1–6.
8. Swayne DE. (2004) Application of new vaccine technologies for the control of transboundary diseases. *Dev Biol (Basel)* **119**: 219–228.
9. Szmuness W, Stevens CE, Harley EJ, *et al.* (1980) Hepatitis B vaccine: demonstration of efficacy in a controlled clinical trial in a high-risk population in the United States. *N Engl J Med* **303**(15): 833–841.
10. Wood JM, Robertson JS. (2004) From lethal virus to life-saving vaccine: developing inactivated vaccines for pandemic influenza. *Nat Rev Microbiol* **2**(10): 842–847.
11. Wu JJ, Huang DB, Pang KR, Tyring SK. (2004) Vaccines and immunotherapies for the prevention of infectious diseases having cutaneous manifestations. *J Am Acad Dermatol* **50**(4): 495–528.
12. Bukhari NI, Aziz MT, Jamshaid M. (2005) Comparative immunogenicity of a genetically derived and a plasma pooled hepatitis B vaccine in normal adult volunteers. *Therapie* **60**(3): 311–316.
13. Yap SF. (2004) Hepatitis B: review of development from the discovery of the "Australia Antigen" to end of the twentieth Century. *Malays J Pathol* **26**(1): 1–12.

14. Belshe RB, Nichol KL, Black SB, *et al.* (2004) Safety, efficacy, and effectiveness of live, attenuated, cold-adapted influenza vaccine in an indicated population aged 5-49 years. *Clin Infect Dis* **39**(7): 920–927.
15. Cox MM. (2005) Cell-based protein vaccines for influenza. *Curr Opin Mol Ther* **7**(1): 24–29.
16. Daemer RJ, Feinstone SM, Gust ID, Purcell RH. (1981) Propagation of human hepatitis A virus in African green monkey kidney cell culture: primary isolation and serial passage. *Infect Immun* **32**(1): 388–393.
17. Flehmig B. (1980) Hepatitis A-virus in cell culture: I. propagation of different hepatitis A-virus isolates in a fetal rhesus monkey kidney cell line (Frhk-4). *Med Microbiol Immunol (Berl)* **168**(4): 239–248.
18. Provost PJ, Hilleman MR. (1979) Propagation of human hepatitis A virus in cell culture *in vitro*. *Proc Soc Exp Biol Med* **160**(2): 213–221.
19. Barbara JA. (2004) The rationale for pathogen-inactivation treatment of blood components. *Int J Hematol* **80**(4): 311–316.
20. Caceres VM, Sutter RW. (2001) Sabin monovalent oral polio vaccines: review of past experiences and their potential use after polio eradication. *Clin Infect Dis* **33**(4): 531–541.
21. Colgrove J, Bayer R. (2005) Could it happen here? Vaccine risk controversies and the specter of derailment. *Health Aff (Millwood)* **24**(3): 729–739.
22. Faure E. (2005) Multiple sclerosis and hepatitis B vaccination: could minute contamination of the vaccine by partial hepatitis B virus polymerase play a role through molecular mimicry? *Med Hypotheses* **65**(3): 509–520.
23. Huang DB, Wu JJ, Tyring SK. (2004) A review of licensed viral vaccines, some of their safety concerns, and the advances in the development of investigational viral vaccines. *J Infect* **49**(3): 179–209.
24. Magrath D, Reeve P. (1993) On the role of the World Health Organization in the development of Sabin vaccines. *Biologicals* **21**(4): 345–348.
25. O'hagan DT, Rappuoli R. (2004) The safety of vaccines. *Drug Discov Today* **9**(19): 846–854.
26. Seligman SJ, Gould EA. (2004) Live flavivirus vaccines: reasons for caution. *Lancet* **363**(9426): 2073–2075.
27. Radetsky P. (1991) *The Invisible Invaders: Viruses and the Scientists Who Pursue Them*. Little, Brown and Company, Boston, New York, Toronto and London.

28. Emini EA, Ellis RW, Miller WJ, *et al.* (1986) Production and immunological analysis of recombinant hepatitis B vaccine. *J Infect* **13** Suppl A: 3–9.
29. Howard CR, Young PR, Lee S, *et al.* (1986) Hepatitis B surface antigen polypeptide micelles from antigen expressed in Saccharomyces cerevisiae. *J Virol Methods* **14**(1): 25–35.
30. Keating GM, Noble S. (2003) Recombinant hepatitis B vaccine (Engerix-B): a review of its immunogenicity and protective efficacy against hepatitis B. *Drugs* **63**(10): 1021–1051.
31. McAleer WJ, Buynak EB, Maigetter RZ, *et al.* (1984) Human hepatitis B vaccine from recombinant yeast. *Nature* **307**(5947): 178–180.
32. Valenzuela P, Medina A, Rutter WJ, *et al.* (1982) Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature* **298**(5872): 347–350.
33. Waters JA, Bailey C, Love C, Thomas HC. (1998) A study of the antigenicity and immunogenicity of a new hepatitis B vaccine using a panel of monoclonal antibodies. *J Med Virol* **54**(1): 1–6.
34. Beyene A, Basu A, Meyer K, Ray R. (2002) Hepatitis C virus envelope glycoproteins and potential for vaccine development. *Vox Sang* **83** Suppl 1: 27–32.
35. Bhopale GM, Nanda RK. (2004) Prospects for hepatitis C vaccine. *Acta Virol* **48**(4): 215–221.
36. Hurwitz JL, Slobod KS, Lockey TD, *et al.* (2005) Application of the polyvalent approach to HIV-1 vaccine development. *Curr Drug Targets Infect Disord* **5**(2): 143–156.
37. Idemyor V. (2003) Human immunodeficiency virus: scientific challenges impeding candidate vaccines. *HIV Clin Trials* **4**(6): 421–424.
38. Lechmann M, Liang TJ. (2000) Vaccine development for hepatitis C. *Semin Liver Dis* **20**(2): 211–226.
39. Logvinoff C, Major ME, Oldach D, *et al.* (2004) Neutralizing antibody response during acute and chronic hepatitis C virus infection. *Proc Natl Acad Sci USA* **101**(27): 10149–10154.
40. Nara PL, Lin G. (2005) HIV-1: the confounding variables of virus neutralization. *Curr Drug Targets Infect Disord* **5**(2): 157–170.
41. Barany G, Kneib-Cordonier N, Mullen DG. (1987) Solid-phase peptide synthesis: a silver anniversary report. *Int J Pept Protein Res* **30**(6): 705–739.
42. Lebl M, Hachmann J. (2005) High-throughput peptide synthesis. *Meth Mol Biol* **298**: 167–194.

43. Merrifield RB, Stewart JM. (1965) Automated peptide synthesis. *Nature* **207**(996): 522–523.
44. Pon RT. (2002) Chemical Synthesis of Oligonucleotides: From Dreams to Automation. In: Y. Khudyakov & H. Fields (eds), *Artificial DNA: Methods and Applications*, pp. 1–70. CRC Press LLC, Boca Raton, Florida.
45. Pumpens P, Grens E. (2002) Artificial genes for chimeric virus-like particles. In: Y. Khudyakov & H. Fields (eds). *Artificial DNA: Methods and Applications*, pp. 249–327. CRC Press LLC, Boca Raton, Florida.
46. Anderer FA. (1963) Preparation and properties of an artificial antigen immunologically related to tobacco mosaic virus. *Biochim Biophys Acta* **71**: 246–248.
47. Langbeheim H, Arnon R, Sela M. (1976) Antiviral effect on MS-2 coliphage obtained with a synthetic antigen. *Proc Natl Acad Sci USA* **73**(12): 4636–4640.
48. Brusic V, Bajic VB, Petrovsky N. (2004) Computational methods for prediction of T-cell epitopes — a framework for modelling, testing, and applications. *Methods* **34**(4): 436–443.
49. Flower DR. (2003) Towards in silico prediction of immunogenic epitopes. *Trends Immunol* **24**(12): 667–674.
50. Carter JM. (1994) Epitope mapping of a protein using the Geysen (PEPSCAN) procedure. *Methods Mol Biol* **36**: 207–223.
51. Toyoda T, Masunaga K, Ohtsu Y, *et al.* (2000) Antibody-scanning and epitope-tagging methods; molecular mapping of proteins using antibodies. *Curr Protein Pept Sci* **1**(3): 303–308.
52. Alix AJ. (1999) Predictive estimation of protein linear epitopes by using the program PEOPLE. *Vaccine* **18**(3–4): 311–314.
53. Thornton JM, Edwards MS, Taylor WR, Barlow DJ. (1986) Location of 'continuous' antigenic determinants in the protruding regions of proteins. *EMBO J* **5**(2): 409–413.
54. Agterberg M, Adriaanse H, Lankhof H, *et al.* (1990) Outer membrane PhoE protein of Escherichia coli as a carrier for foreign antigenic determinants: immunogenicity of epitopes of foot-and-mouth disease virus. *Vaccine* **8**(1): 85–91.
55. Brown AL, Francis MJ, Hastings GZ, *et al.* (1991) Foreign epitopes in immunodominant regions of hepatitis B core particles are highly immunogenic and conformationally restricted. *Vaccine* **9**(8): 595–601.
56. Colbere-Garapin F, Christodoulou C, Crainic R, *et al.* (1988) Addition of a foreign oligopeptide to the major capsid protein of poliovirus. *Proc Natl Acad Sci USA* **85**(22): 8668–8672.

57. Mastico RA, Talbot SJ, Stockley PG. (1993) Multiple presentation of foreign peptides on the surface of an RNA-free spherical bacteriophage capsid. *J Gen Virol* **74** (Pt 4): 541–548.
58. Stocker BA, Newton SM. (1994) Immune responses to epitopes inserted in Salmonella flagellin. *Int Rev Immunol* **11**(2): 167–178.
59. Sverdlov ED, Tsarev SA, Markova SV, *et al.* (1989) Insertion of short hepatitis virus A amino acid sequences into poliovirus antigenic determinants results in viable progeny. *FEBS Lett* **257**(2): 354–356.
60. Boarino A, Scalone A, Gradoni L, *et al.* (2005) Development of recombinant chimeric antigen expressing immunodominant B epitopes of Leishmania infantum for serodiagnosis of visceral leishmaniasis. *Clin Diagn Lab Immunol* **12**(5): 647–653.
61. Chang JC, Ruedinger B, Cong M, *et al.* (1999) Artificial NS4 mosaic antigen of hepatitis C virus. *J Med Virol* **59**(4): 437–450.
62. Khudyakov YE, Favorov MO, Khudyakova NS, *et al.* (1994) Artificial mosaic protein containing antigenic epitopes of hepatitis E virus. *J Virol* **68**(11): 7067–7074.
63. Lin S, Arcangel P, Medina-Selby A, *et al.* (2005) Design of novel conformational and genotype-specific antigens for improving sensitivity of immunoassays for hepatitis C virus-specific antibodies. *J Clin Microbiol* **43**(8): 3917–3924.
64. Liu W, Peng Z, Liu Z, *et al.* (2004) High epitope density in a single recombinant protein molecule of the extracellular domain of influenza A virus M2 protein significantly enhances protective immunity. *Vaccine* **23**(3): 366–371.
65. Dow BC, Buchanan I, Munro H, *et al.* (1996) Relevance of RIBA-3 supplementary test to HCV PCR positivity and genotypes for HCV confirmation of blood donors. *J Med Virol* **49**(2): 132–136.
66. Fujinami RS, Oldstone MB. (1989) Molecular mimicry as a mechanism for virus-induced autoimmunity. *Immunol Res* **8**(1): 3–15.
67. Ionescu RM, Przysiecki CT, Liang X, *et al.* (2006) Pharmaceutical and immunological evaluation of human papillomavirus virus-like particle as an antigen carrier. *J Pharm Sci* **95**(1): 70–79.
68. Khudyakov Y, Favorov MO, Jue DL, *et al.* (1994) Immunodominant antigenic regions in a structural protein of the hepatitis E virus. *Virology* **198**(1): 390–393.
69. Srinivasappa J, Saegusa J, Prabhakar BS, *et al.* (1986) Molecular mimicry: frequency of reactivity of monoclonal antiviral antibodies with normal tissues. *J Virol* **57**(1): 397–401.

70. Trujillo JR, McLane MF, Lee TH, Essex M. (1993) Molecular mimicry between the human immunodeficiency virus type 1 gp120 V3 loop and human brain proteins. *J Virol* **67**(12): 7711–7715.
71. Galarza JM, Latham T, Cupo A. (2005) Virus-like particle vaccine conferred complete protection against a lethal influenza virus challenge. *Viral Immunol* **18**(2): 365–372.
72. Jeong SH, Qiao M, Nascimbeni M, *et al.* (2004) Immunization with hepatitis C virus-like particles induces humoral and cellular immune responses in nonhuman primates. *J Virol* **78**(13): 6995–7003.
73. Matsuo E, Tani H, Lim C, *et al.* (2006) Characterization of HCV-like particles produced in a human hepatoma cell line by a recombinant baculovirus. *Biochem Biophys Res Commun* **340**(1): 200–208.
74. Edman JC, Hallewell RA, Valenzuela P, *et al.* (1981) Synthesis of hepatitis B surface and core antigens in *E. coli*. *Nature* **291**(5815): 503–506.
75. Gough NM, Murray K. (1982) Expression of the hepatitis B virus surface, core and E antigen genes by stable rat and mouse cell lines. *J Mol Biol* **162**(1): 43–67.
76. Hardy K, Stahl S, Kupper H. (1981) Production in *B. subtilis* of hepatitis B core antigen and a major antigen of foot and mouth disease virus. *Nature* **293**(5832): 481–483.
77. Nemeckova S, Sroller V, Kunke D, *et al.* (1996) Hepatitis B virus corepreS2 particles expressed by recombinant vaccinia virus. *Acta Virol* **40**(5–6): 273–279.
78. Pumpens P, Borisova GP, Crowther RA, Grens E. (1995) Hepatitis B virus core particles as epitope carriers. *Intervirology* **38**(1–2): 63–74.
79. Schodel F, Kelly S, Tinge S, *et al.* (1996) Hybrid hepatitis B virus core antigen as a vaccine carrier moiety. II. Expression in avirulent Salmonella spp. for mucosal immunization. *Adv Exp Med Biol* **397**: 15–21.
80. Seifer M, Hamatake R, Bifano M, Standring DN. (1998) Generation of replication-competent hepatitis B virus nucleocapsids in insect cells. *J Virol* **72**(4): 2765–2776.
81. Standring DN, Ou JH, Masiarz FR, Rutter WJ. (1988) A signal peptide encoded within the precore region of hepatitis B virus directs the secretion of a heterogeneous population of e antigens in Xenopus oocytes. *Proc Natl Acad Sci USA* **85**(22): 8405–8409.
82. Tsuda S, Yoshioka K, Tanaka T, *et al.* (1998) Application of the human hepatitis B virus core antigen from transgenic tobacco plants for serological diagnosis. *Vox Sang* **74**(3): 148–155.

83. Borisova GP, Berzins I, Pushko PM, *et al.* (1989) Recombinant core particles of hepatitis B virus exposing foreign antigenic determinants on their surface. *FEBS Lett* **259**(1): 121–124.
84. Schodel F, Peterson D, Hughes J, Milich DR. (1993) A virulent Salmonella expressing hybrid hepatitis B virus core/pre-S genes for oral vaccination. *Vaccine* **11**(2): 143–148.
85. Schodel F, Peterson D, Hughes J, *et al.* (1996) Hybrid hepatitis B virus core antigen as a vaccine carrier moiety: I. presentation of foreign epitopes. *J Biotechnol* **44**(1–3): 91–96.
86. Stahl SJ, Murray K. (1989) Immunogenicity of peptide fusions to hepatitis B virus core antigen. *Proc Natl Acad Sci USA* **86**(16): 6283–6287.
87. Grene E, Mezule G, Borisova G, *et al.* (1997) Relationship between antigenicity and immunogenicity of chimeric hepatitis B virus core particles carrying HIV type 1 epitopes. *AIDS Res Hum Retroviruses* **13**(1): 41–51.
88. Isaguliants MG, Nordlund S, Sallberg M, *et al.* (1996) HIV-1 epitopes exposed by hybrid hepatitis B core particles affect proliferation of peripheral blood mononuclear cells from HIV-1 positive donors. *Immunol Lett* **52**(1): 37–44.
89. Geysen HM, Rodda SJ, Mason TJ. (1986) *A priori* delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol Immunol* **23**(7): 709–715.
90. Geysen HM, Rodda SJ, Mason TJ, *et al.* (1987) Strategies for epitope analysis using peptide synthesis. *J Immunol Methods* **102**(2): 259–274.
91. Lam KS, Salmon SE, Hersh EM, *et al.* (1991) A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **354**(6348): 82–84.
92. Tujeta R. (2002) Mimotopes: an overview. In: Y. Khudyakov & H. Fields (eds), *Artificial DNA: Methods and Applications*, pp. 225–247. CRC Press LLC, Boca Raton, Florida.
93. Felici F, Luzzago A, Folgori A, Cortese R. (1993) Mimicking of discontinuous epitopes by phage-displayed peptides, II. Selection of clones recognized by a protective monoclonal antibody against the Bordetella pertussis toxin from phage peptide libraries. *Gene* **128**(1): 21–27.
94. Luzzago A, Felici F, Tramontano A, *et al.* (1993) Mimicking of discontinuous epitopes by phage-displayed peptides, I. Epitope mapping of human H ferritin using a phage library of constrained peptides. *Gene* **128**(1): 51–57.

95. Moreau V, Granier C, Villard S, *et al.* (2006) Discontinuous epitope prediction based on mimotope analysis. *Bioinformatics* **22**(9): 1088–1095.
96. Scott JK. (1992) Discovering peptide ligands using epitope libraries. *Trends Biochem Sci* **17**(7): 241–245.
97. Felici F, Castagnoli L, Musacchio A, *et al.* (1991) Selection of antibody ligands from a large library of oligopeptides expressed on a multivalent exposition vector. *J Mol Biol* **222**(2): 301–310.
98. Bellofiore P, Petronzelli F, De Martino T, *et al.* (2006) Identification and refinement of a peptide affinity ligand with unique specificity for a monoclonal anti-tenascin-C antibody by screening of a phage display library. *J Chromatogr A* **1107**(1–2): 182–191.
99. Dou XG, Talekar G, Chang J, *et al.* (2002) Antigenic heterogeneity of the hepatitis C virus NS5A protein. *J Clin Microbiol* **40**(**1**): 61–67.
100. Biles BD, Connolly BA. (2004) Low-fidelity Pyrococcus furiosus DNA polymerase mutants useful in error-prone PCR. *Nucleic Acids Res* **32**(22): e176.
101. Neylon C. (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res* **32**(4): 1448–1459.
102. Willats WG. (2002) Phage display: practicalities and prospects. *Plant Mol Biol* **50**(6): 837–854.
103. Schooltink H, Rose-John S. (2005) Designing cytokine variants by phage-display. *Comb Chem High Throughput Screen* **8**(2): 173–179.
104. Rothe A, Hosse RJ, Power BE. (2006) Ribosome display for improved biotherapeutic molecules. *Expert Opin Biol Ther* **6**(2): 177–187.
105. Partidos CD, Chirinos-Rojas CL, Steward MW. (1997) The potential of combinatorial peptide libraries for the identification of inhibitors of TNF-alpha mediated cytotoxicity *in vitro*. *Immunol Lett* **57**(1–3): 113–116.
106. Walden P, Wiesmuller KH, Jung G. (1995) Elucidation of T-cell epitopes: a synthetic approach with random peptide libraries. *Biochem Soc Trans* **23**(3): 678–681.
107. Harayama S. (1998) Artificial evolution by DNA shuffling. *Trends Biotechnol* **16**(2): 76–82.
108. Whalen RG, Kaiwar R, Soong NW, Punnonen J. (2001) DNA shuffling and vaccines. *Curr Opin Mol Ther* **3**(1): 31–36.
109. Locher CP, Paidhungat M, Whalen RG, Punnonen J. (2005) DNA shuffling and screening strategies for improving vaccine efficacy. *DNA Cell Biol* **24**(4): 256–263.

110. Arbib MA. (2003) The Elements of Brain Theory and Neural Networks. In: M.A. Arbib (ed), *The Handbook of Brain Theory and Neural Networks*, pp. 3–23. The MIT Press, Cambridge, Massachusetts.
111. Fausett L. (1994) *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications.* Prentice Hall, Upper Saddle River, New Jersey.
112. Kecman V. (2001) Basic Tools of Soft Computing: Neural Networks, Fuzzy Logic Systems, and Support Vector Machines. In: *Learning and Soft Computing: Support Vector Machines, Neural Networks and Fuzzy Logic Models*, pp. 13–18. The MIT Press, Cambridge, Massachusetts.
113. Hagan MT, Demuth HB, Beale M. (1996) Introduction. In: *Neural Network Design*, pp. 1–12. PWS Publishing Company, Boston, Massachusetts.
114. Haykin S. (1999) Introduction. In: *Neural Networks*, pp. 1–49. Prentice Hall, Upper Saddle River, New Jersey.
115. Wu CH, McLarty JW. (2000) *Neural Networks and Genome Informatics*, First Edition. Elsevier Science, Kidlington, Oxford.
116. Casari G, Sander C, Valencia A. (1995) A method to predict functional residues in proteins. *Nat Struct Biol* **2**(2): 171–178.
117. Principe JC, Euliano NR, Lefebvre WC. (2000) *Neural and Adaptive Systems: Fundamentals Through Simulations.* John Wiley & Sons, Inc., New York.
118. Zavaljevski N, Stevens FJ, Reifman J. (2002) Support vector machines with selective kernel scaling for protein classification and identification of key amino acid positions. *Bioinformatics* **18**(5): 689–696.
119. Brusic V, Rudy G, Harrison L. (1995) Prediction of MHC binding peptides using artificial neural networks. *Complexity International* **2**(April): 1–9.
120. Schneider G, Wrede P. (1998) Artificial neural networks for computer-based molecular design. *Prog Biophys Mol Biol* **70**(3): 175–222.
121. Edelman J. (1993) Quadratic minimization of predictors for protein secondary structure. Application to transmembrane alpha-helices. *J Mol Biol* **232**(1): 165–191.
122. Eisenberg D, Schwarz E, Komaromy M, Wall R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* **179**(1): 125–142.
123. Engelman DM, Steitz TA, Goldman A. (1986) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu Rev Biophys Biophys Chem* **15**: 321–353.

124. Juretic D, Lucic B, Trinajstic N. (1993) Predicting membrane protein secondary structure. Preference functions method for finding optimal conformational preferences. *Croatica Chemica Acta* **66**: 201–208.
125. Kyte J, Doolittle RF. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**(1): 105–132.
126. O'Neil KT, DeGrado WF. (1990) A Thermodynamic Scale for the Helix-Forming Tendencies of the Commonly Occurring Amino Acids. *Science* **250**(4981): 646–651.
127. Grantham R. (1974) Amino acid difference formula to help explain protein evolution. *Science* **185**(4154): 862–864.
128. Woese CR, Dugre DH, Dugre SA, *et al.* (1966) On the fundamental nature and evolution of the genetic code. *Cold Spring Harb Symp Quant Biol* **31**: 723–736.
129. Mathusamy R, Ponnuswamy PK. (1990) Variation of amino properties in protein secondary structures, alpha-helices and beta-strands. *Int J Pept Protein Res* **35**(5): 378–395.
130. Jones DD. (1975) Amino Acid Properties and Side-chain Orientation in Proteins. *J Theor Biol* **21**(1): 167–183.
131. Chou PY. (1989) Prediction of Protein Structural Classes from Amino Acid Composition. In: G.D. Fasman (ed), *Prediction of Protein Structure and the Principles of Protein Conformation*, pp. 549–586. Plenum, New York.
132. Hopp TP, Woods KR. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA* **78**(6): 3824–3828.
133. Fauchere JL, Charton M, Kier LB, *et al.* (1988) Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int J Pept Protein Res* **32**(4): 269–278.
134. Gasteiger E, Hoogland C, Gattiker A, *et al.* (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: J.M. Walker (ed), *The Proteomics Protocols Handbook*, pp. 571-607. Humana Press, Herts, UK.
135. Wang J. (1997) Feature Selection Using Neural Networks with Contribution Measures. *IEEE Transactions on Neural Networks* **8**(4): 645–662.
136. Leray P, Gallinary P. (1999) Feature Selection with Neural Networks. *Behaviormetrika* **26**(1): 145–166.
137. Heckerling PS, Gerber BS, Tape TG, Wigton RS. (2005) Selection of predictor variables for pneumonia using neural networks and genetic algorithms. *Meth Inf Med* **44**(1): 89–97.

138. Heckerling PS, Gerber BS, Tape TG, Wigton RS. (2003) Entering the black box of neural networks. *Methods Inf Med* **42**(3): 287–296.
139. Holbrook SR, Muskal SM, Kim SH. (1990) Predicting surface exposure of amino acids from protein sequence 4. *Protein Eng* **3**(8): 659–665.
140. Eisenberg D. (1984) Three-dimensional structure of membrane and surface proteins. *Annu Rev Biochem* **53**: 595–623.
141. Calise AJ, Lee S, Sharma M. (2001) Development of a Reconfigurable Flight Control Law for Tailless Aircraft. *AIAA J Guid, Control Dyn* **24**(5): 896–902.
142. Johnson EN, Proctor AA, Ha J. (2004) Development and test of highly autonomous unmanned aerial vehicles. *AIAA J Aerospace Comput, Info Comm* **1**(1)(January): 5–18.
143. Miglino O, Walker R. (2002) Genetic redundancy in evolving populations of simulated robots. *Artif Life* **8**(3): 265–277.
144. Ritter H, Steil JJ, Nolker C, *et al.* (2003) Neural architectures for robot intelligence. Rev Neurosci **14**(1–2): 121–143.
145. Ramesh AN, Kambhampati C, Monson JR, Drew PJ. (2004) Artificial intelligence in medicine. *Ann R Coll Surg Engl* **86**(5): 334–338.
146. Traeger M, Eberhart A, Geldner G, *et al.* (2003) Artificial neural networks. Theory and applications in anesthesia, intensive care and emergency medicine. *Anaesthesist* **52**(11): 1055–1061.
147. Walczak S. (2001) An Empirical Analysis of Data Requirements for Financial Forecasting with Neural Networks. *J Manage Info Syst* **17**(4): 203–222.
148. Maqsood I, Khan MR, Huang GH, Abdalla R. (2005) Application of soft computing models to hourly weather analysis in southern Saskatchewan, Canada. *Eng Appl Artif Intell* **18**(1): 115–125.
149. So SS, Karplus M. (1996) Evolutionary optimization in quantitative structure-activity relationship: an application of genetic neural networks. *J Med Chem* **39**(7): 1521–1530.
150. Zhao CY, Zhang RS, Zhang HX, *et al.* (2005) QSAR study of natural, synthetic and environmental endocrine disrupting compounds for binding to the androgen receptor. *SAR QSAR Environ Res* **16**(4): 349–367.
151. Sutherland JJ, O'Brien LA, Weaver DF. (2004) A comparison of methods for modeling quantitative structure-activity relationships. *J Med Chem* **47**(22): 5541–5554.

152. Eric S, Solmajer T, Zupan J, *et al.* Prediction of selectivity of alpha1-adrenergic antagonists by counterpropagation neural network (CP-ANN). *Farmaco* **59**(5): 389–395.
153. Douali L, Villemin D, Zyad A, Cherqaoui D. (2004) Artificial neural networks: non-linear QSAR studies of HEPT derivatives as HIV-1 reverse transcriptase inhibitors. *Mol Divers* **8**(1): 1–8.
154. Bazoui H, Zahouily M, Boulajaaj S, *et al.* (2002) QSAR for anti-HIV activity of HEPT derivatives. *SAR QSAR Environ Res* **13**(6): 567–577.
155. Davis SJ, Ikemizu S, Evans EJ, *et al.* (2003) The nature of molecular recognition by T cells. *Nat Immunol* **4**(3): 217–224.
156. van der Merwe PA, Davis SJ. (2003) Molecular interactions mediating T cell antigen recognition. *Annu Rev Immunol* **21**: 659–684.
157. Goldsby RA, Kindt TJ, Osborne BA. (2000) Major Histocompatibility Complex. In J. Kuby (ed), *Immunology*, pp. 173–197. W.H. Freeman and Company, New York.
158. Robinson J, Malik A, Parham P, *et al.* (2000) IMGT/HLA database — a sequence database for the human major histocompatibility complex. *Tiss Antig* **55**(3): 280–287.
159. Robinson J, Waller MJ, Parham P, *et al.* (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucl Acids Res* **31**(1): 311–314.
160. Bankovich AJ, Girvin AT, Moesta AK, Garcia KC. (2004) Peptide register shifting within the MHC groove: theory becomes reality. *Mol Immunol* **40**(14–15): 1033–1039.
161. Brusic V, Bajic VB, Petrovsky N. (2004) Computational methods for prediction of T-cell epitopes — a framework for modelling, testing, and applications. *Methods* **34**(4): 436–443.
162. Buus S. (1999) Description and prediction of peptide-MHC binding: the human MHC project. *Curr Opin Immunol* **11**(2): 209–213.
163. Honeyman MC, Brusic V, Stone NL, Harrison LC. (1998) Neural network-based prediction of candidate T-cell epitopes. *Nat Biotechnol* **16**(10): 966–969.
164. Berzofsky JA, Ahlers JD, Belyakov IM. (2001) Strategies for Designing and Optimizing New Generation Vaccines. *Nat Rev Imminol* **1**(3): 209–219.
165. Brusic V, Rudy G, Harrison LC. (1998) MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res* **26**(1): 368–371.
166. McSparron H, Blythe MJ, Zygouri C, *et al.* (2003) JenPep: a novel computational information resource for immunobiology and vaccinology. *J Chem Inf Comput Sci* **43**(4): 1276–1287.

167. Robinson J, Waller MJ, Parham P, *et al.* (2001) IMGT/HLA — a sequence database for the human major histocompatibility complex. *Nucl Acids Res* **29**(1): 210–213.
168. Rammensee H, Bachmann J, Emmerich NP, *et al.* (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**(3–4): 213–219.
169. Bhasin M, Singh H, Raghava GP. (2003) MHCBN: a comprehensive database of MHC binding and non-binding peptides. *Bioinformatics* **19**(5): 665–666.
170. Bhasin M, Raghava GP. (2004) Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine* **22**(23–24): 3195–3204.
171. DeLisi C, Berzofsky JA. (1985) T-cell antigenic sites tend to be amphipathic structures. *Proc Natl Acad Sci USA* **82**(20): 7048–7052.
172. Pincus MR, Gerewitz F, Schwartz RH, Scheraga HA. (1983) Correlation between the conformation of cytochrome c peptides and their stimulatory activity in a T-lymphocyte proliferation assay. *Proc Natl Acad Sci USA* **80**(11): 3297–3300.
173. Spouge JL, Guy HR, Cornette JL, *et al.* (1987) Strong conformational propensities enhance T cell antigenicity. *J Immunol* **138**(**1**): 204–212.
174. Reyes VE, Fowlie EJ, Lu S, *et al.* (1990) Comparison of three related methods to select T cell-presented sequences of protein antigens. *Mol Immunol* **27**(10): 1021–1027.
175. Margalit H, Spouge JL, Cornette JL, *et al.* (1987) Prediction of immunodominant helper T cell antigenic sites from the primary sequence. *J Immunol* **138**(7): 2213–2229.
176. Stille CJ, Thomas LJ, Reyes VE, Humphreys RE. (1987) Hydrophobic strip-of-helix algorithm for selection of T cell-presented peptides. *Mol Immunol* **24**(10): 1021–1027.
177. Rothbard JB, Taylor WR. (1988) A sequence pattern common to T cell epitopes. *EMBO J* **7**(1): 93–100.
178. Mouritsen S, Meldal M, Ruud-Hansen J, Werdelin O. (1991) T-helper-cell determinants in protein antigens are preferentially located in cysteine-rich antigen segments resistant to proteolytic cleavage by cathepsin B, L, and D. *Scand J Immunol* **34**(4): 421–431.
179. Meister GE, Roberts CG, Berzofsky JA, De Groot AS. (1995) Two novel T cell epitope prediction algorithms based on MHC-binding motifs; comparison of predicted and published epitopes from Mycobacterium tuberculosis and HIV protein sequences. *Vaccine* **13**(6): 581–591.

180. Deavin AJ, Auton TR, Greaney PJ. (1996) Statistical comparison of established T-cell epitope predictors against a large database of human and murine antigens. *Mol Immunol* **33**(2): 145–155.
181. Falk K, Rotzschke O, Stevanovic S, *et al.* (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**(6324): 290–296.
182. Guillet JG, Hoebeke J, Lengagne R, *et al.* (1991) Haplotype specific homology scanning algorithm to predict T-cell epitopes from protein sequences. *J Mol Recognit* **4**(1): 17–25.
183. Nijman HW, Houbiers JG, Vierboom MP, *et al.* (1993) Identification of peptide sequences that potentially trigger HLA-A2.1-restricted cytotoxic T lymphocytes. *Eur J Immunol* **23**(6): 1215–1219.
184. Sette A, Buus S, Appella E, *et al.* (1989) Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proc Natl Acad Sci USA* **86**(9): 3296–3300.
185. Hattotuwagama CK, Doytchinova IA, Flower DR. (2005) In silico prediction of peptide binding affinity to class I mouse major histocompatibility complexes: a comparative molecular similarity index analysis (CoMSIA) study. *J Chem Inf Model* **45**(5): 1415–1423.
186. Bordner AJ, Abagyan R. (2006) Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins* **63**(3): 512–526.
187. Gulukota K, Sidney J, Sette A, DeLisi C. (1997) Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J Mol Biol* **267**(5): 1258–1267.
188. Kast WM, Brandt RM, Sidney J, *et al.* (1994) Role of HLA-A motifs in identification of potential CTL epitopes in human papillomavirus type 16 E6 and E7 proteins. *J Immunol* **152**(8): 3904–3912.
189. Rammensee HG, Friede T, Stevanoviic S. (1995) MHC ligands and peptide motifs: first listing. *Immunogenetics* **41**(4): 178–228.
190. Sidney J, Oseroff C, del Guercio MF, *et al.* (1994) Definition of a DQ3.1-specific binding motif. *J Immunol* **152**(9): 4516–4525.
191. Brusic V, Schonbach C, Takiguchi M, *et al.* (1997) Application of genetic search in derivation of matrix models of peptide binding to MHC molecules. *Proc Int Conf Intell Syst Mol Biol* **5**: 75–83.
192. Hammer J, Bono E, Gallazzi F, *et al.* (1994) Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. *J Exp Med* **180**(6): 2353–2358.

193. Kondo A, Sidney J, Southwood S, *et al.* (1995) Prominent roles of secondary anchor residues in peptide binding to HLA-A24 human class I molecules. *J Immunol* **155**(9): 4307–4312.
194. Mallios RR. (2001) Predicting class II MHC/peptide multi-level binding with an iterative stepwise discriminant analysis meta-algorithm. *Bioinformatics* **17**(10): 942–948.
195. Parker KC, Bednarek MA, Coligan JE. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J Immunol* **152**(1): 163–175.
196. Rothbard JB, Marshall K, Wilson KJ, *et al.* (1994) Prediction of peptide affinity to HLA DRB1*0401. *Int Arch Allergy Immunol* **105**(1): 1–7.
197. Schafer JR, Jesdale BM, George JA, *et al.* (1998) Prediction of well-conserved HIV-1 ligands using a matrix-based algorithm, EpiMatrix. *Vaccine* **16**(19): 1880–1884.
198. Schonbach C, Kun Y, Brusic V. (2002) Large-scale computational identification of HIV T-cell epitopes. *Immunol Cell Biol* **80**(3): 300–306.
199. Udaka K, Wiesmuller KH, Kienle S, *et al.* (2000) An automated prediction of MHC class I-binding peptides based on positional scanning with peptide libraries. *Immunogenetics* **51**(10): 816–828.
200. Brusic V, Rudy G, Honeyman G, *et al.* (1998) Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics* **14**(2): 121–130.
201. Milik M, Sauer D, Brunmark AP, *et al.* (1998) Application of an artificial neural network to predict specific class I MHC binding peptide sequences. *Nat Biotechnol* **16**(8): 753–756.
202. Nielsen M, Lundegaard C, Worning P, *et al.* (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* **12**(5): 1007–1017.
203. Borras-Cuesta F, Golvano J, Garcia-Granero M, *et al.* (2000) Specific and general HLA-DR binding motifs: comparison of algorithms. *Hum Immunol* **61**(3): 266–278.
204. Adams HP, Koziol JA. (1995) Prediction of binding to MHC class I molecules. *J Immunol Methods* **185**(2): 181–190.
205. Jackson MR, Song ES, Yang Y, Peterson PA. (1992) Empty and peptide-containing conformers of class I major histocompatibility complex molecules expressed in Drosophila melanogaster cells. *Proc Natl Acad Sci USA* **89**(24): 12117–12121.

206. Matsumura M, Saito Y, Jackson MR, *et al.* (1992) *In vitro* peptide binding to soluble empty class I major histocompatibility complex molecules isolated from transfected Drosophila melanogaster cells. *J Biol Chem* **267**(33): 23589–23595.
207. Hammer J, Takacs B, Sinigaglia F. (1992) Identification of a motif for HLA-DR1 binding peptides using M13 display libraries. *J Exp Med* **176**(4): 1007–1013.
208. Jameson SC, Bevan J. (1992) Dissection of major histocompatibility complex (MHC) and T cell receptor contact residues in a Kb-restricted ovalbumin peptide and assessment of the predictive power of MHC-binding motifs. *Eur J Immunol* **22**: 2663–2667.
209. Taylor WR. (1986) The classification of amino acid conservation. *J Theor Biol* **119**(2): 205–218.
210. Hammer J, Nagy ZA, Sinigaglia F. (1994) Rules governing peptide-class II MHC molecule interactions. *Behring Inst Mitt* **94**: 124–132.
211. Hammer J, Bono E, Givehchi A, *et al.* (1994) Precise prediction of MHC class II-peptide interaction based on peptide side chain scanning. *J Exp Med* **180**: 2353–2358.
212. Schonbach C, Ibe M, Shiga H, *et al.* (1995) Fine tuning of peptide binding to HLA-B*3501 molecules by nonanchor residues. *J Immunol* **154**(11): 5951–5958.
213. Southwood S, Sidney J, Kondo A, *et al.* (1998) Several common HLA-DR types share largely overlapping peptide binding repertoires. *J Immunol* **160**(7): 3363–3373.
214. Honeyman MC, Brusic V, Harrison LC. (1997) Strategies for identifying and predicting islet autoantigen T-cell epitopes in insulin-dependent diabetes mellitus. *Ann Med* **29**(5): 401–404.
215. Nielsen M, Lundegaard C, Lund O, Kesmir C. (2005) The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **57**(1–2): 33–41.
216. Zhang GL, Khan AM, Srinivasan KN, *et al.* (2005) MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides. *Nucl Acids Res* **33**(Web Server issue): W172–W179.
217. Devillers J. (1996) *Neural Networks in QSAR and Drug Design*. Academic Press, San Diego.
218. Lohmann R, Schneider G, Wrede P. (1996) Structure optimization of an artificial neural filter detecting membrane-spanning amino acid sequences. *Biopolymers* **38**(1): 13–29.

219. Schneider G, Schuchhardt J, Wrede P. (1994) Artificial neural networks and simulated molecular evolution are potential tools for sequence-oriented protein design. *Comput Appl Biosci* **10**(6): 635–645.
220. Schneider G, Schuchhardt J, Wrede P. (1995) Development of simple fitness landscapes for peptides by artificial neural filter systems. *Biol Cybern* **73**(3): 245–254.
221. Wrede P, Landt O, Klages S, *et al.* (1998) Peptide design aided by neural networks: biological activity of artificial signal peptidase I cleavage sites. *Biochemistry* **37**(11): 3588–3593.
222. Xiao Y, Segal MR. (2005) Prediction of Genomewide Conserved Epitope Profiles of HIV-1: Classifier Choice and Peptide Representation. *Stat Appl Genet Mol Biol* **4**(1): Article 25.
223. Zhao Y, Pinilla C, Valmori D, *et al.* (2003) Application of support vector machines for T-cell epitopes prediction. *Bioinformatics* **19**(15): 1978–1984.
224. Corbet S, Nielsen HV, Vinner L, *et al.* (2003) Optimization and immune recognition of multiple novel conserved HLA-A2, human immunodeficiency virus type 1-specific CTL epitopes. *J Gen Virol* **84**(Pt 9): 2409–2421.
225. Lauemoller SL, Kesmir C, Corbet SL, *et al.* (2000) Identifying cytotoxic T cell epitopes from genomic and proteomic information: "The human MHC project". *Rev Immunogenet* **2**(4): 477–491.
226. Gombar VK, Silver IS, Zhao Z. (2003) Role of ADME characteristics in drug discovery and their *in silico* evaluation: *in silico* screening of chemicals for their metabolic stability. *Curr Top Med Chem* **3**(11): 1205–1225.
227. Schneider G, Schrodl W, Wallukat G, *et al.* (1998) Peptide design by artificial neural networks and computer-based evolutionary search. *Proc Natl Acad Sci USA* **95**(21): 12179–12184.
228. Harpaz Y, Gerstein M, Chothia C. (1994) Volume changes on protein folding. *Structure* **2**(7): 641–649.
229. Eigen M, McCaskill J, Schuster P. (1988) Molecular quasi-species. *J Phys Chem* **92**(24): 6881–6891.
230. Schneider G, Fechner U. (2005) Computer-Based De Novo Design of Drug-Like Molecules. *Nature Rev Drug Discovery* **4**(8): 649–663.
231. Flower DR, McSparron H, Blythe MJ, *et al.* (2003) Computational vaccinology: quantitative approaches. *Novartis Found Symp* **254**: 102–120.
232. Doytchinova IA, Flower DR. (2002) Quantitative approaches to computational vaccinology. *Immunol Cell Biol* **80**(3): 270–279.

233. Wang L, Brock A, Herberich B, Schultz PG. (2001) Expanding the genetic code of Escherichia coli. *Science* **292**(5516): 498–500.
234. Wang L, Schultz PG. (2002) Expanding the genetic code. *Chem Commun (Camb)* **1**: 1–11.
235. Wang L, Schultz PG. (2004) Expanding the genetic code. *Angew Chem Int Ed Engl* **44**(1): 34–66.
236. Deiters A, Geierstanger BH, Schultz PG. (2005) Site-specific *in vivo* labeling of proteins for NMR studies. *Chembiochem* **6**(1): 55–58.
237. Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG. (1989) A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**(4901): 182–188.
238. Xie J, Wang L, Wu N, *et al.* (2004) The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat Biotechnol* **22**(10): 1297–1301.
239. Xu R, Hanson SR, Zhang Z, *et al.* (2004) Site-specific incorporation of the mucin-type N-acetylgalactosamine-alpha-O-threonine into protein in Escherichia coli. *J Am Chem Soc* **126**(48): 15654–15655.
240. Hohsaka T, Ashizuka Y, Taira H, *et al.* (2001) Incorporation of nonnatural amino acids into proteins by using various four-base codons in an Escherichia coli *in vitro* translation system. *Biochemistry* **40**(37): 11060–11064.
241. Hohsaka T, Fukushima M, Sisido M. (2002) Nonnatural mutagenesis in E. coli and rabbit reticulocyte lysates by using four-base codons. *Nucl Acids Res Suppl* **2**: 201–202.
242. Hohsaka T, Sisido M. (2002) Incorporation of non-natural amino acids into proteins. *Curr Opin Chem Biol* **6**(6): 809–815.
243. Muranaka N, Hohsaka T, Sisido M. (2006) Four-base codon mediated mRNA display to construct peptide libraries that contain multiple nonnatural amino acids. *Nucl Acids Res* **34**(1): e7.
244. Hohsaka T, Sisido M. (2000) Incorporation of nonnatural amino acids into proteins by using five-base codon-anticodon pairs. *Nucl Acids Symp Ser* **44**: 99–100.
245. Hohsaka T, Ashizuka Y, Murakami H, Sisido M. (2001) Five-base codons for incorporation of nonnatural amino acids into proteins. *Nucl Acids Res* **29**(17): 3646–3651.
246. Krebs S, Folkers G, Rognan D. (1998) Binding of rationally designed non-natural peptides to the human leukocyte antigen HLA-B*2705. *J Pept Sci* **4**(6): 378–388.

247. Klein J, Perciva T, Stäring J. (2000) Prediction of MHC II-peptide binding using genetic programming. (Unpublished data; J Klein, personal communication (JK@artificial.com)).
248. Gillet VJ. (2000) *De Novo* Molecular Design. In D.E. Clark (ed), *Evolutionary Algorithms in Molecular Design*, pp. 49–66. Wiley-VCH, New York.
249. Koza JR. (1998) *Genetic Programming II*. The MIT Press, Cambridge, Massachusetts.
250. Hiltzik M., A.I. Reboots. Technology Review.com [MAR 2002], 1–4. 2005. MIT Enterprise. Ref Type: Electronic Citation.
251. Chinnasamy A, Sung WK, Mittal A. (2004) Protein structure and fold prediction using tree-augmented naive Bayesian classifier. *Pac Symp Biocomput* **9**: 387–398.
252. Klingler TM, Brutlag DL. (1994) Discovering side-chain correlation in alpha-helices. *Proc Int Conf Intell Syst Mol Biol* **2**: 236–243.
253. Robles V, Larranaga P, Pena JM, *et al.* (2004) Bayesian network multiclassifiers for protein secondary structure prediction. *Artif Intell Med* **31**(2): 117–136.
254. Jansen R, Yu H, Greenbaum D, *et al.* (2003) A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* **302**(5644): 449–453.
255. Lin N, Wu B, Jansen R, *et al.* (2004) Information assessment on predicting protein-protein interactions. *BMC Bioinformatics* **5**(1): 154.
256. Nariai N, Kim S, Imoto S, Miyano S. (2004) Using protein-protein interactions for refining gene networks estimated from microarray data by Bayesian networks. *Pac Symp Biocomput* **9**: 336–347.
257. Chang JH, Hwang KB, Oh SJ, Zhang BT. (2005) Bayesian network learning with feature abstraction for gene-drug dependency analysis. *J Bioinform Comput Biol* **3**(1): 61–77.
258. Ben Gal I, Shani A, Gohr A, *et al.* (2005) Identification of transcription factor binding sites with variable-order Bayesian networks. *Bioinformatics.*
259. Chen TM, Lu CC, Li WH. (2005) Prediction of splice sites with dependency graphs and their expanded Bayesian networks. *Bioinformatics* **21**(4): 471–482.
260. Beal MJ, Falciani F, Ghahramani Z, *et al.* (2005) A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics* **21**(3): 349–356.

261. Missal K, Cross MA, Drasdo D. (2006) Gene network inference from incomplete expression data: transcriptional control of hematopoietic commitment. *Bioinformatics* **22**(6): 731–738.
262. Nariai N, Tamada Y, Imoto S, Miyano S. (2005) Estimating gene regulatory networks and protein-protein interactions of Saccharomyces cerevisiae from multiple genome-wide data. *Bioinformatics* **21**(2): ii206–ii212.
263. Perrin BE, Ralaivola L, Mazurie A, *et al.* (2003) Gene networks inference using dynamic Bayesian networks. *Bioinformatics* **19**(2): II138–II148.
264. Burnside ES. (2005) Bayesian networks Computer-assisted diagnosis support in radiology(1). *Acad Radiol* **12**(4): 422–430.
265. Kahn CE, Jr., Laur JJ, Carrera GF. (2001) A Bayesian network for diagnosis of primary bone tumors. *J Digit Imaging* **14**(2 Suppl 1): 56–57.
266. Charniak E. (1991) Bayesian Networks without Tears. *AI Magazine*, **12**(4): 50–63. American Association for Artificial Intelligence (AAAI) Press.
267. Pearl J. (2000) *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York.
268. Korb KB, Nicholson AE. (2004) *Bayesian Artificial Intelligence*. CRC Press LLC, Boca Raton, Florida.
269. Neapolitan RE. (2004) *Learning Bayesian Networks*, First Edition. Pearson Prentice Hall, Upper Saddle River, New Jersey.
270. Pearl J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Second Edition. Morgan Kaufman Publishers, San Francisco, California.
271. Astakhov V, Cherkasov A. (2005) Prediction of HLA-A2 binding peptides using Bayesian Networks. *Bioinformation* **1**(2): 58–63.
272. Kohonen T. (2001) *Self-Organizing Maps*, Third Edition. Springer-Verlag, Berlin.
273. Alahakoon D, Halgamuge SK, Srinivasan B. (2000) Dynamic Self-Organizing Maps with Controlled Growth for Knowledge Discovery. *IEEE Trans on Neural Networks* **2**(3): 601–614.
274. Fritzke B. (1994) Growing cell structures — a self-organizing network for unsupervised and supervised learning. *Neural Networks* **7**(9): 1441–1460.

Chapter 10

Bioinformatics Resources for the Study of Viruses at the Virginia Bioinformatics Institute

Anjan Purkayastha[†], *Oswald Crasta*[†], *J Dana Eckart*[‡], *Michael Czar*[*,†], *XJ Meng*[§], *Joao Setubal*[†] *and Bruno Sobral*[†]

Introduction

This chapter gives an overview of two bioinformatics resources that are being developed at the Virginia Bioinformatics Institute (VBI) for the study of pathogenic microorganisms, with a special focus on pathogens listed as biodefense priority agents by the Centers for Disease Control (CDC) and the National Institute of Allergy and Infectious Diseases (NIAID). The Pathosystems Resource Integration Center (PATRIC) has been formed to provide a comprehensive and centralized annotation resource for eight classes of human pathogens, including all sequenced species, subspecies, strains and isolates. The Pathogen Portal (PathPort) was established to provide the scientific community with an integrated biological data analysis platform that includes multiple and diverse data sets on microbial pathogens and a set of analysis tools.

*Corresponding author.
[†]Virginia Bioinoformatics Institute, Virginia Polytechnic and State University, Blacksburg, VA, USA.
[‡]AFLAC, Columbus, GA, USA.
[§]VA-MD Regional College of Veterinary Medicine, Virginia Polytechnic and State University, Blacksburg, VA, USA.

Advances in genomics and computer technologies are fueling the development of a knowledgebase from large data sets created by high-throughput genomic technologies using digital computation, data, information, and networks. The integration of knowledge from various fields such as computer science, chemistry, and biology has created a vast opportunity by creating new research environments based upon Cyberinfrastructure.[1] Cyberinfrastructure refers to the infrastructure required to develop a knowledgebase based upon distributed computing, information, and communication technology. One of the major challenges during the post-genomic era is the integration of diverse data sets.[2] As described by Kanehisa and Bork, the main goal of bioinformatics during the 1990s was to create primary databases of genes and proteins.[3] Currently the focus is to extend the databases for quantitative data from transcriptomes and proteomes and to provide interoperability between multiple disciplines. The bioinformatics resources developed under such endeavors are expected to assist the scientific community in making use of the large datasets produced by high-throughput technologies in elucidation of the mechanisms of phenotypic expression. Understanding of these mechanisms may result in the identification of specific targets and products relevant to public health. Both PATRIC and PathPort fall under the auspices of the Cyberinfrastructure Group at VBI, whose goal is to provide a scalable, flexible and interoperable technology framework that facilitates biodiscovery.

Bioinformatics Resource Centers

In July 2004, NIAID funded eight centers to serve the needs of scientists investigating select groups of pathogens, which include many viruses. These centers will provide comprehensive genomics resources on pathogens causing emerging or re-emerging diseases or considered to be priorities for biodefense. We describe in this section the general goals of NIAID's Bioinformatics Resource Centers (BRC) initiative.

NIAID strives to understand, treat, and ultimately prevent infectious, immunologic, and allergic diseases that threaten millions of human lives. Within NIAID, the Division of Microbiology and Infectious Diseases (DMID) provides grants for research to control and prevent diseases

caused by virtually all infectious agents with the exception of AIDS. The programs supported range from basic biomedical research, which includes studies of microbial physiology and antigenic structure, to applied research, which includes the development of diagnostic tests, to clinical trials for evaluating experimental drugs and vaccines. Since the terrorist events in September of 2001, a number of initiatives have been launched with a focus on infectious agents that are relevant to biodefense. NIAID has already made a significant investment in genome sequencing of invertebrate vectors of disease and pathogenic microorganisms, including medically important microbes, and those considered agents of bioterrorism. Coupled with other biochemical and microbiological information, these sequences are facilitating the development of targets for diagnostics, drugs, and vaccines, and for forensic strain identification. Vast amounts of information have been and are currently being generated, and the storage, integration, and interpretation of this information requires a cyberinfrastructure consisting of relational databases and robust software components and tools to facilitate data exploration by the scientific community.

A NIAID initiative to meet these needs of integration and analysis is the Bioinformatics Resource Centers for Biodefense and Emerging or Re-Emerging Infectious Diseases (BRC). Eight BRC centers were established, under an initial five-year grant, to construct a cyberinfrastructure that integrates genomic, proteomic, biochemical, and microbiological information (http://www.niaid.nih.gov/dmid/genomes/brc/default.htm). A list of the eight BRCs and their affiliated institutions is given in Table 1. The BRC initiative will provide the scientific community with a research resource that includes access to a large amount of genomic and related data, and to tools that facilitate the interpretation of that data. The BRC resources are for public use and are be offered completely free of charge.

The specific objectives of this initiative are:

1. To provide the scientific community with a robust point of entry for access to genomic and related data in a user-friendly format. The centers focus on the organisms listed in the NIAID Category A-C priority list or pathogens causing emerging or re-emerging diseases that are relevant for biodefense.

Table 1. List of the eight Bioinformatics Resource Centers funded by NIAID in 2004

	BRC Name	Institutes	Link
1	PathoSystems Resource Integration Center (PATRIC)	Virginia Bioinformatics Institute (VBI) Loyola University Medical Center Social and Scientific Systems University of Maryland	http://patric.vbi.vt.edu/
2	Pathema	The J. Craig Venter Institute	http://www.tigr.org/pathema/
3	Aplicomplexan Database Resources (ApiDB)	University of Pennsylvania University of Georgia	http://apidb.org/
4	Vector Base	University of Notre Dame (UND) European Bioinformatics Institute European Molecular Biology Laboratory Institute of Molecular Biology and Biotechnology Harvard University Purdue University University of California Riverside	http://www.vectorbase.org/

(*Continued*)

Table 1. (*Continued*)

	BRC Name	Institutes	Link
5	National Microbial Pathogen Data Resource (NMPDR)	University of Chicago The Fellowship for Interpretation of Genomes (FIG) Argonne National Laboratory The University of Illinois	http://www.nmpdr.org/
6	Viral Bioinformatics Resource Center (VBRC)	University of Alabama Birmingham (UAB) University of Victoria, Canada	http://www.biovirus.org/
7	The Biodefense/Public Health DataBase (BioHealthBase)	University of Texas Southwestern Medical Center (UTSMC) Northrop Grumman Vecna Technologies Amar International	http://www.biohealthbase.org/
8	Enteropathogen Resource Integration Center (ERIC)	SRA International University of Wisconsin Madison	http://www.ericbrc.org/eric/

2. To facilitate the identification of potential targets for the development of vaccines, therapeutics and diagnostics.

The eight BRCs will curate genomic information on a specific set of microorganisms and integrate other types of data relevant to those pathogens. The list of the pathogens in the NIAID Category A to C priority and the corresponding BRC associated are shown in Fig. 1. A mandate for interoperability between BRCs has been made by NIAID, and standardization of data formats will be critical to meet this goal. Each BRC is developing a multi-organism relational database that it will maintain and continually update. Standard and advanced analytical tools will also be built and enhanced throughout the duration of the project as a companion to the database and an analysis resource to help researchers access and analyze the data. For example, the analysis tools will enable comparative analysis of genomes to allow the identification of genetic polymorphisms that correlate with phenotypes such as pathogenicity, drug resistance, morbidity and infectivity. A total of three out of the eight BRCs (PATRIC, VBRC and BioHealthBase) (Table 1) are involved in the curation of the viruses. Here we describe the organization of the resources at PATRIC.

The Pathogen Resource Integration Center (PATRIC) and Annotation and Curation of Microbial Genomic Data

PATRIC will annotate and curate genomic and related biological data for eight pathogenic microorganisms, which include three bacteria and five viruses. The bacteria are *Brucella*, *Rickettsia*, and *Coxiella*, while the viruses include the caliciviruses, coronaviruses, hepatitis A, hepatitis E, and rabies viruses. Table 2 provides the number of genomes stored at PATRIC and the range of genome sizes. Specifically, we plan to annotate and curate biological information on the multiple strains/genotypes of the hepatitis A virus, the hepatitis E virus and the rabies virus species; members of the genus *coronavirus* (family *Coronaviridae*); and the entire *Caliciviridae* family.

Systems biology is the study of biological organisms at the systems-level that takes into account the interactions of the key cellular components (DNA, RNA, protein, metabolic pathways) and the response of

Category A

- Bacillus anthracis (anthrax) [2]
- Clostridium botulinum [2]
- Yersinia pestis [8]
- Variola major (smallpox) & other pox viruses [6]
- Francisella tularensis (tularemia) [7]
- Viral hemorrhagic fevers [6]
  - Arenaviruses
    - LCM, Junin virus, Machupo virus, Guanarito virus
    - Lassa Fever
  - Bunyaviruses
    - Hantaviruses
    - Rift Valley Fever
  - Flaviruses
    - Dengue
  - Filoviruses
    - Ebola
    - Marburg

Category B

- Burkholderia pseudomallei [2]
- Coxiella burnetii (Q fever) [1]
- Brucella species (brucellosis) [1]
- Burkholderia mallei (glanders) [2]
- Ricin toxin (from Ricinus communis) [7]
- Epsilon toxin of Clostridium perfringens [2]
- Staphylococcus enterotoxin B [5]
- Typhus fever (Rickettsia prowazekii) [1]
- Food and Waterborne Pathogens
  - Bacteria
    - Diarrheagenic E.coli [8]
    - Pathogenic Vibrios [5]
    - Shigella species [8]
    - Salmonella [8]
    - Listeria monocytogenes [5]
    - Campylobacter jejuni [5]
    - Yersinia enterocolitica [8]
  - Viruses (Caliciviruses, Hepatitis A) [1]
  - Protozoa
    - Cryptosporidium parvum [3]
    - Cyclospora cayatanensis
    - Giardia lamblia [7]
    - Entamoeba histolytica [2]
    - Toxoplasma [3]
    - Microsporidia [7]
- Additional viral encephalitides
  - West Nile Virus [6]
  - LaCrosse
  - California encephalitis [6]
  - VEE [6]
  - EEE [6]
  - WEE [6]
  - Japanese Encephalitis Virus [6]
  - Kyasanur Forest Virus [6]

Category C Emerging infectious disease threats such as Nipah virus[6] and additional hantaviruses.

NIAID priority areas:

- Tickborne hemorrhagic fever viruses [6]
  - Crimean-Congo Hemorrhagic fever virus
- Tickborne encephalitis viruses [6]
- Yellow fever [6]
- Multi-drug resistant TB [7]
- Influenza [7]
- Other Rickettsias [1]
- Rabies [1]
- Severe acute respiratory syndrome-associated coronavirus (SARS-CoV) [1] *(Note: SARS-CoV added August 30, 2004)*

Fig. 1. The list of pathogens in the Categories A, B and C as assigned by NIAID. (Source: http://www2.niaid.nih.gov/biodefense/bandc_priority.htm). The number following the pathogen name indicates the corresponding BRC assigned as in Table 1.

cells to their environment. At PATRIC, we define the pathogen-host-environment triangle as a PathoSystem. Different factors at each of these nodes converge to influence the outcome of an infection, which may range from microbe-clearance to full-blown disease. It will be our

Table 2. Genomes stored at PATRIC. The three bacterial and five viral organisms are listed along with the number of genomes stored at PATRIC (* as of July 2007) and the range of genome sizes for each organism

	Organism Expert *(Affiliation)*	Number of Genomes (whole and partial)*	Genome Size
Bacteria			
Brucella	Dr. Steven Boyle *(Virginia Tech)*	6	Ch1: 2.11 Mb Ch2: 1.17 Mb
Coxiella	Dr. Robert Heinzen *(NIAID, NIH)* Dr. James Samuel *(Texas A&M University)*	4	1.5–2.1 Mb
Rickettsiae	Dr. Abdu Azad *(University of Maryland)*	12	1.1–1.6 Mb
Viruses			
Coronaviruses	Dr. Susan Baker *(Loyola University of Chicago)*	232	27–31 kb
Caliciviruses	Dr. Jan Vinje *(Centers for Disease Control and Prevention)*	77	7.7–8.5 kb
Hepatitis A Viruses	Dr. Yuri Khudyakov *(Centers for Disease Control and Prevention)*	48	7.4 kb
Hepatitis E Viruses	Dr. Xiang-Jin Meng *(Virginia Tech)*	65	7.2 kb
Rabies Viruses	Dr. Charles Rupprecht *(Centers for Disease Control and Prevention)*	30	11.9 kb

aim to curate information on the entire PathoSystem and not just the pathogen. We believe that this holisitic, "systems biology" approach to studying a disease will yield valuable insights into the pathogenic mechanisms of these microbes, knowledge that will be crucial in the development of novel diagnostics, vaccines, and therapeutics.

Central to the concept of systems biology is the integration of theory, wet lab experimentation, modeling, and simulation in the context of any biological process. In following this approach we will strive to provide the scientific community with an integrated view of biological data drawn from numerous sources. The PATRIC database

will contain high-quality annotated data: sequence annotations from published whole and partial genomes; relevant experimental data; metabolic pathway data; taxonomic and ontological data; relevant literature, and a suite of visualization and analysis tools. An organism expert (OE) — a scientist with several years of experience and a recognized authority in the field — will help direct the curation of each organism and connect PATRIC to the specific research community (see Table 2).

Genome Curation Process

The main goal of the curation process is to provide well-curated resources for the organisms to the scientific community. The genome curation process is depicted in Fig. 2 and is classified into following main categories:

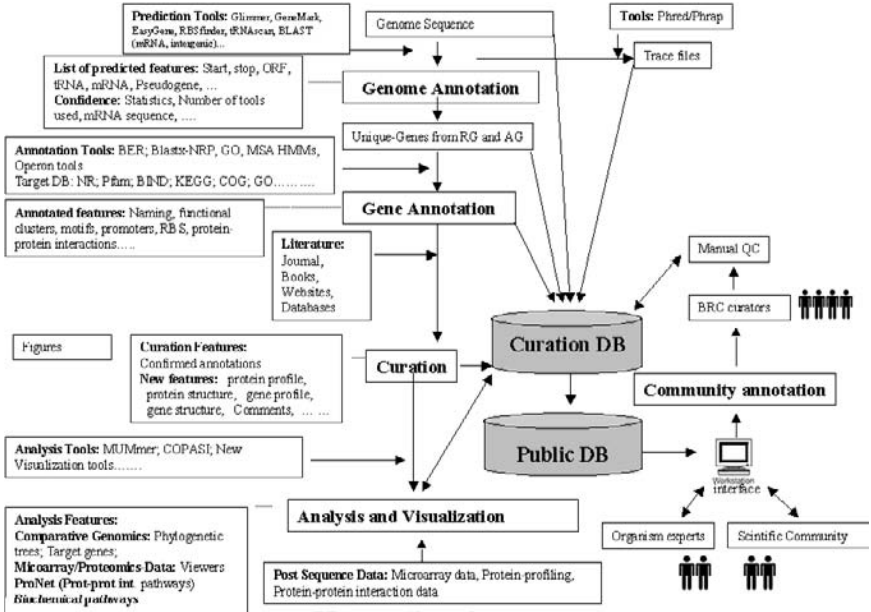

Prediction Tools: Glimmer, GeneMark, EasyGene, RBSfinder, tRNAscan, BLAST (mRNA, intergenic)...
Genome Sequence
Tools: Phred/Phrap
Trace files
List of predicted features: Start, stop, ORF, tRNA, mRNA, Pseudogene, ...
Confidence: Statistics, Number of tools used, mRNA sequence, ...
Genome Annotation
Unique-Genes from RG and AG
Annotation Tools: BER; Blastx-NRP, GO, MSA HMMs, Operon tools
Target DB: NR; Pfam; BIND; KEGG; COG; GO.....
Gene Annotation
Annotated features: Naming, functional clusters, motifs, promoters, RBS, protein-protein interactions.....
Literature: Journal, Books, Websites, Databases
Manual QC
BRC curators
Figures
Curation Features: Confirmed annotations
New features: protein profile, protein structure, gene profile, gene structure, Comments, ...
Curation
Curation DB
Community annotation
Public DB
Analysis Tools: MUMmer; COPASI; New Visualization tools.....
Workstation interface
Analysis Features:
Comparative Genomics: Phylogenetic trees; Target genes;
Microarray/Proteomics Data: Viewers
ProNet (Prot-prot int. pathways)
Biochemical pathways
.....
Analysis and Visualization
Organism experts
Scientific Community
Post Sequence Data: Microarray data, Protein-profiling, Protein-protein interaction data

Fig. 2. The PATRIC curation schema. Displayed is the flow of information from the genomic sequence of an organism to other types of systemic data: transcriptomic, proteomic, and metabolomic. Also shown are the various individuals who will take part in the annotation process.

1. Genome annotation: identification of features present in the genomic sequence.
2. Gene annotation: analysis of each protein-coding gene.
3. Literature-based genome curation: integration of literature information relevant to genes or the genome.
4. Post-sequence curation: integration of data from experimental sources.
5. Comparative genomics: comparisons at three different levels: 1. whole genome alignments, 2. identification of orthologs at the gene level, and 3. identification of SNPs and other polymorphisms at nucleotide/residue level.

Pathogen Background Information

Besides the genome curation, PATRIC's curation team will generate detailed background information on pathogens using the XML-based Pathogen Information Markup Language (PIML) as described by He *et al.*[4] PIML allows for a portable, system-independent, machine-parseable, and human-readable representation of general information for any pathogen. This body of pathogen information can be queried and displayed graphically from a web service (http://staff.vbi.vt.edu/pathport/services/wsdls/piml.wsdl) in ToolBus/PathPort,[5] which provides a custom graphical visualization module for PIML documents and interoperability with other types of infectious disease data (e.g. genomic). A web-based query and display system was also developed to query the complete pathogen information or a specific topic across multiple pathogens (http://www.vbi.vt.edu/pathport/pathinfo/query.html). These documents are expected to serve the scientific community with up-to-date information on various aspects of the pathogens with the listing of the corresponding bibliography.

Schema for Curation of Biological Data

A significant challenge in designing PATRIC is the storage and retrieval of diverse kinds of biological data. Our computational infrastructure must be robust, scalable and flexible to permit the integration of data on different organisms from multiple sources and allow

easy access for computational analyses. We have chosen the Genomics Unified Schema (GUS) database architecture (http://www.gusdb.org) to address our information storage needs. GUS offers a schema and user interface that integrates genomic, transcriptomic and proteomic data of multiple organisms, along with information on their metabolic and regulatory networks. It also has provisions to store ontologies and controlled vocabularies, gene expression data, and comparative genomic data. GUS is an open-source, relational database schema that has been developed by the Computational Biology and Informatics Laboratory at the University of Pennsylvania. It has been successfully implemented by other biological web-resources. GUS currently uses the Oracle database management system, and there are plans to port it to other DBMSs.

Genomic Analysis Tools

PATRIC's software systems will play three primary roles: support the in-house genomic annotation/curation efforts; provide the scientific community with access to the curated information, together with an evolving suite of analysis tools aimed at helping researchers develop vaccines, therapeutics, and diagnostics; and enable consumers of this information to provide the project curators with both general and specific feedback. Users will be able to browse the archived genome annotations via powerful query and navigation capabilities. The various levels of biological information will be seamlessly interconnected so that a user may move, for example, from viewing gene annotations to querying metabolic pathways to viewing results of high-throughput experiments and a summary of published literature. The analysis tools will support both whole-genome and gene-level comparisons.

The Scientific Community — Outreach and Feedback

In addition to providing information, PATRIC will engage the scientific communities it serves through a series of outreach activities. These activities include establishing and maintaining contact with genome providers (sequencing centers), participating in scientific meetings relevant to the organisms we curate, and seeking and accepting

feedback from the community. We invite and welcome participation of the community at all levels of our efforts, from helping us better curate specific information of specific genomes to suggestions of tools that can help the community better take advantage of the wealth of information available on our website. There are specific mechanisms for feedback at our website (http://patric.vbi.vt.edu), and comments can also be sent to the e-mail address patric@vbi.vt.edu.

Viruses Curated at PATRIC

This section offers an introduction to each of the five viral PathoSystems that are being curated at PATRIC. Current knowledge of the taxonomy, structure, and host-range of each PathoSystem are briefly described.

The Genus Coronavirus

The genus *coronavirus* is a member of the family *Coronaviridae*.[6] Coronaviruses are large, enveloped, single-stranded, positive-sense RNA viruses. With genome sizes ranging from 27 to 32 kB, the coronaviruses have the largest genomes of all the known RNA viruses. They have a spherical lipid-bilayer envelope enclosing a nucleocapsid core. The lipid envelope has multiple copies of the integral M (membrane) and S (spike) glycoproteins, the latter protruding outwards in the form of club-shaped spikes. Some coronaviruses (belonging to Group 2, see below) possess a third integral membrane glycoprotein: HE (hemagglutinin-esterase) protein. The genomic RNA molecule, which is capped and polyadenylated, associates with copies of the N (nucleocapsid) protein to form a long, helical nucleocapsid. A typical coronavirus genome has two large, overlapping open-reading frames (ORF), ORF1a and ORF1b, that encode two polyproteins, and a set of ORFs that encode structural (S, M, N) and other accessory proteins. While all coronaviruses encode the two polyproteins, the number and composition of the other ORFs vary between species.

Coronaviruses are found in birds and mammals. The viruses fall into three categories, based on serological cross-reactivity:

Groups 1 and 2 infect mammalian hosts while Group 3 infects only avian hosts.

In animals, coronavirus infections can lead to virulent multi-systemic diseases. For example, in pigs, the porcine transmissible gastroenteritis virus (TGEV) causes respiratory and enteric infections; in mice, the mouse hepatitis virus (MHV) causes respiratory, enteric and neurologic infections as well as hepatitis.[6]

In humans, coronaviruses were associated with mild diseases until the appearance of the SARS coronavirus, which causes severe, often life-threatening, disease. Most human coronaviruses cause a mild, upper respiratory tract infection. They are transmitted via the aerosol route and direct as well as indirect contact. The incubation period of the virus is about three days, leading to the symptoms of a common cold: fever, runny nose, cough and malaise. The disease is resolved within a few days. An exceptional case is the SARS virus, a newly discovered human coronavirus.[7] The outcome of SARS infection varies from a mild illness to a more severe disease leading to respiratory failure and death. Phylogenetic analysis suggests that the SARS virus has split off from the Group 2 coronaviruses.[7]

The Caliciviridae Family

The *Caliciviridae* family is comprised of viruses that are known to cause disease in humans and other animal species.[8] The family is divided into four genera: *Vesivirus*, *Lagovirus*, *Norovirus*, and *Sapovirus*.[9] Caliciviruses have a single-stranded, non-segmented, positive-sense RNA genome of approximately 7.7–8.5 kB. Family members encode two or three ORFs, the arrangements of which differ between genera. The viruses have a non-enveloped, icosahedral capsid composed of a single protein type. The Norwalk virus (genus *Norovirus*) capsid protein has two major domains, S and P. The S domain forms the inner shell while the P domain forms arch-like structures that extend outward from the shell.[10]

Caliciviruses infect different species and cause remarkably different diseases in their hosts.[9] The ability to correlate the type of disease caused by the virus to features in its genomic sequence is something that the PATRIC project aims to enable. In swine, the vesicular

exanthema virus (VESV) (genus *Vesivirus*) causes vesicles on the snout and hooves. In cats, the feline calicivirus (FCV) (genus *Vesivirus*) causes conjunctivitis, ulcers, limping, and respiratory disease. The *Lagovirus* species cause extensive hepatic necrosis in rabbits and hares. In humans, the *Noroviruses* and the *Sapoviruses* are transmitted primarily by the fecal-oral route and are the most important causes of epidemic nonbacterial gastroenteritis. The virus has a mean incubation period of about 24 hours, with the illness lasting from 24 to 48 hours.[8] Norwalk virus gastroenteritis is characterized by nausea, vomiting and diarrhea; the disease is usually self-limiting.

Hepatitis A Virus

The hepatitis A virus (HAV) is the only member of the genus *Hepatovirus* (*Picornaviridae* family).[11] The HAV genome is a single-stranded, non-segmented, positive-sense RNA molecule of size 7.4 kB. It contains just one gene that encodes a large polyprotein of 2227 amino acids that is proteolytically cleaved to form ten smaller, functional proteins. The genomic RNA molecule is uncapped but polyadenylated, and it has a viral protein (Vpg) covalently attached to its 5′ end. HAV has a non-enveloped, dodecahedral capsid that is composed of four types of structural polypeptides. The genomic RNA molecule, along with the covalently linked Vpg, forms the inner core of the virion.

The virus causes disease in humans and several primate species. Transmission occurs mainly via the fecal-oral route by the ingestion of food or water contaminated with infected feces. The course of HAV infection is highly variable, ranging from asymptomatic infection to acute disease. The ability to correlate symptoms of the infection to features in its genomic sequence is something that the PATRIC project aims to enable. The disease has four progressive stages: incubation, prodromal, icteric, and convalescent. The incubation period ranges from 10 to 50 days, followed by fatigue, fever, nausea and vomiting. In the icteric stage, abnormally high bilirubin levels cause a discoloration of the mucous membranes, skin and conjunctivae; the patient discharges dark, golden-brown urine and pale stools. Most patients

recover from the disease; however, fatalities do occur, with case-fatality rates being highest for older adults.

Hepatitis E Virus

The hepatitis E virus (HEV) is the prototype species of the genus *Hepevirus*, the only member of the family *Hepeviridae*.[12] It has a single-stranded, non-segmented, positive-sense RNA genome of about 7.2 kB that encodes three ORFs. Recent work suggests that the HEV capsid is composed of a single protein type.[13] Multiple copies of the structural protein interact to form an icosahedral shell with protrusions.

HEV is known to infect humans and other animal species. It is endemic to various countries. The virus is spread mainly through the fecal-oral route; however, cases of hepatitis E infections arising from blood-transfusions and eating raw/undercooked meat have also been reported.[12,14–16] In humans, the virus causes hepatitis E, an acute disease characterized by anorexia, jaundice and an inflamed liver. It has a low mortality rate of about 1%.

Rabies Virus

The Rabies virus belongs to the *Rhabdoviridae* family, and is the prototypic member of the genus *lyssavirus*.[17] It has a single-stranded, non-segmented, negative-sense RNA genome of approximately 11.9 kB that contains five genes (N, P, G, M and L) and a pseudogene. Like most other members of the *Rhabdoviridae* family, the rabies virus posseses a characterestic "bullet" shape. Each virion has two components: (a) a ribonucleoprotein (RNP) core, which consists of the single-stranded RNA genome tightly coiled around multiple copies of the major nucleoprotein N and the minor polymerase components, the L and P proteins; and (b) a lipoprotein bilayer membrane. Copies of the glycosylated G protein, embedded in the lipid bilayer, protrude outward as spike-like projections. The matrix protein M binds to the inner membrane and to the RNP, forming a bridge between the inner ribonucleoprotein core and the surface membrane. The structural

geometry involved in the formation of the rhabdovirus virion remains to be elucidated.

The virus can infect all warm-blooded animals and causes acute disease of the central nervous system. It is usually transmitted by the bite of infected animals. The disease has five general stages of progression: incubation, prodrome, acute neurologic period, coma, and death.[18] The incubation period — the time between infection and the onset of clinical symptoms — is usually 1 to 2 months; extreme cases in which the incubation period was less than a week or several years have also been reported. The disease is almost always fatal, if not promptly treated.

The Pathogen Portal Project (PathPort)

In addition to providing information from the PATRIC project through a web-browser interface, as required by the contract, we proposed an additional means of access through the Pathogen Portal (PathPort). PathPort, developed at VBI through funding from the Department of Defense, is an integrated platform providing biological data and analysis tools to scientists studying the host-pathogen-environment interaction triangle, with a particular focus on pathogens that are considered to be biodefense priorities. PathPort utilizes a Service Oriented Architecture (SOA) consisting of web-services and a client-side JAVA application, ToolBus, which utilizes plugins for data visualizations. ToolBus acts as a "bus" for contacting data and tools (from the SOA) and serves as a single, consistent user interface. Information on the PathPort project is available at http://pathport.vbi.vt.edu/main/home.php.

Biological data are distributed across different databases (e.g. GenBank, SwissProt) and are stored in a variety of formats. The need for the development of cyberinfrastructure is highlighted by the fact that data analysis tools use different, and often incompatible, data formats. This heterogeneity in the storage, representation, and use of biological data is a challenge that PathPort seeks to address by different data sets and tools. PathPort has been designed to be a flexible, scalable, platform-independent, and seamless system that allows the user access to different kinds of biological data and analysis tools.

As the functioning of ToolBus is not dependent on any data domain type, it can easily support new data types. This allows new analysis and visualization modules to be rapidly added to PathPort. It is scalable both in the number of users and the services it provides, as data sources, analysis tools, and data models are all decentralized and web service tools, are hosted on multiple servers. PathPort uses XML, JAVA and associated technologies, making it a platform-independent application.

The client-side interconnect, ToolBus, provides users with a consistent visualization interface and gives users greater control over data and tools located at off-site servers, allowing them to download only the data necessary for their work. ToolBus also allows users to create groups of associated data; such associations may be made between different data types that reside in separate visualizations. Data association is further supported by a novel function of ToolBus — group suggestors. Group suggestors monitor the data in ToolBus to look for possible associations among the data, which are then presented to the user. Associations can be made on the basis of gene ontology assignments, sequence identifiers (GenBank or Swissprot identifiers), precomputed orthologs (planned), and other shared attributes. ToolBus allows the saving and sharing of work sessions. Because of the platform-independent nature of the application, work sessions may be shared among colleagues working on different operating systems.

PathPort offers a wide range of biological data analysis tools (Table 3). Pathport covers gene prediction programs, sequence alignment (pairwise and multiple), sequence comparison, sequence similarity searches, and microarray data analysis modules. Aside from biological data and analysis tools, PathPort also provides manually curated information in the form of Pathogen Information (PathoInfo) documents and Molecular Interaction Networks (MINet). PathInfo documents contain detailed background information on up to 46 biothreat agents. These manually curated documents describe the taxonomy, lifecycle, epidemiology, and the host range of the pathogen, including any relevant biosafety information, methods of culturing, and diagnostic tests (Fig. 3). PathInfo documents, in the PIML format (described above), may be accessed through ToolBus. The MINet project provides a graphical representation of the knowledge of host-pathogen

Table 3. List of Services at PathPort

Data:	Sequence Alignment and Search:
Genome Search	BLAST
Pathogen Background Information	FASTA
Phylogenetic Tree construction	Smith-Waterman
EBI (gene expression)	CLUSTAL-W
GEO (gene expresion)	Sean
Gene Prediction:	Ssaha and SsahaSNP
Genscan	MUMmer
Glimmer	InterProScan
GlimmerM	Microarray analysis:
TigrScan	Agnes
GrailEXP	Hclust
GeneMark	Kmean
Orpheus	Anova
Probe design:	Manova
PCR/Hybridization	Diana (Hierarchical clustering)
YODA	Multtest (f/t tests)
DNA assembly, digestion:	Rfda (Discriminant analysis)
Contigs from trace files	Rknn (KNN classification)
Restriction enzymes	Rlda (Supervised classification)
tRNA	Rpca (PCA classification)
Transcription factors	Rsom (Gene SOM)
Ribosomal RNA	Rsvm (SVM classification)

interactions gleaned from literature curation (Fig. 4). In order to promote interoperability, PathPort has developed the Molecular Interaction Network Markup Language (MINetML), an XML-based data representation format along the lines of PIML. MINetML focuses on a high-level description of the molecular and cellular interaction between pathogen and host. It emphasizes curatorial accuracy and cross-referencing to external databases and ontology systems.

Outreach Activities at PathPort

The PathPort project is committed to providing the scientific community with integrated and easy access to the latest data and state-of-the-art analysis tools. The project actively solicits feedback and suggestions for improvement from PathPort users. Users may suggest features to

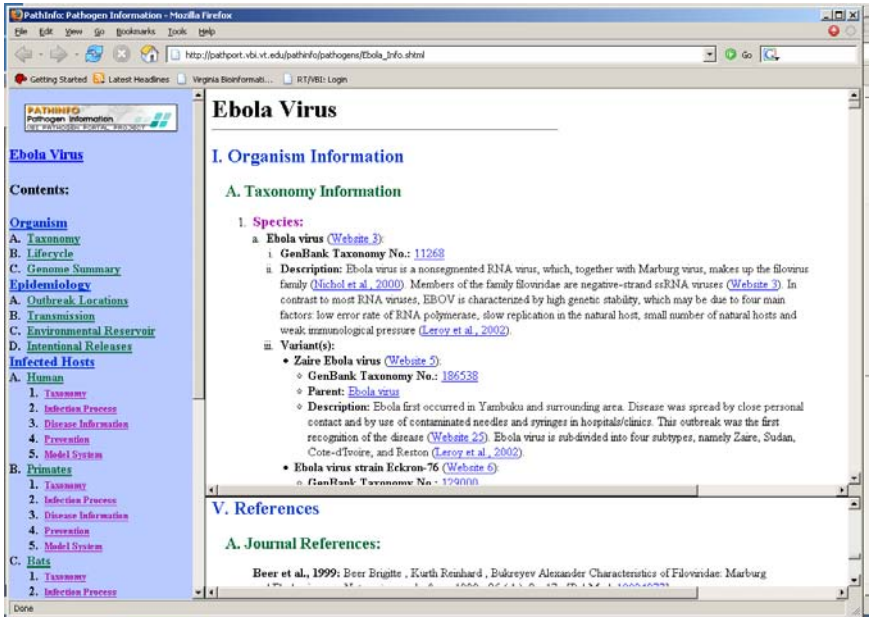


Fig. 3. A sample PathInfo document. PathInfo documents contain manually curated background information on a pathogen. The document is available both in the HTML and PIML formats at the PathPort website (http://pathport.vbi.vt.edu/pathinfo/).

be added at pathport.vbi.vt.edu/scrum. PathPort also holds day-long workshops all over the country to provide hands-on training to members of the scientific community.

Summary and Conclusion

The PATRIC and PathPort projects have been established to facilitate research on the host-pathogen-environment triangle. Both projects have been set up as part of the federal government's initiatives on biodefense. The overarching goal of both PATRIC and PathPort is to provide infectious disease researchers with centralized and integrated portals of information that will facilitate the development of novel therapeutics, vaccines, and diagnostics for known biothreat agents. Both projects have been developed on a flexible infrastructure that

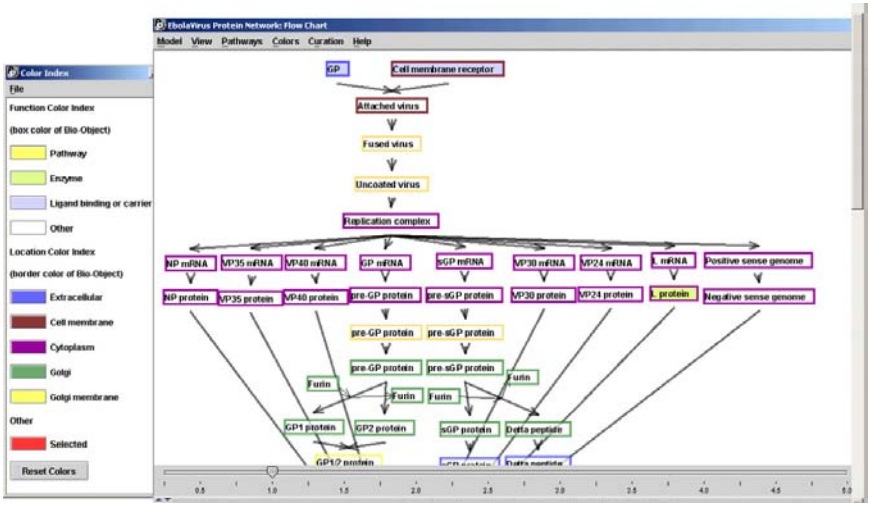


Fig. 4. A sample MINet document available through ToolBus. MINet documents graphically represent the interaction between a pathogen and its host at the cellular/molecular level.

will allow for the rapid inclusion of the most current data and and analysis tools, thus allowing scientists to keep abreast of the latest developments in their areas of research.

References

1. Atkins DE, Droegemeier KK, Feldman SI, *et al.* (2003) Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. http://www.communitytechnology.org/nsf_ci_report/
2. Stein LD. (2003) Integrating biological Databases. *Nature Reviews* **4**: 337–345.
3. Kanehisa M, Bork P. (2003) Bioinformatics in the post-sequence era. *Nature Genet Supplement* **33**: 305–310.
4. He Y, Vines RR, Wattam AR, *et al.* (2005) PIML: the Pathogen Information Markup Language. *Bioinformatics* **21**(1): 116–121.
5. Eckart JD, Sobral BW. (2003) A life scientist's gateway to distributed data management and computing: the PathPort/ToolBus framework. *OMICS* **7**(1): 79–88.

6. Mclntosh K. (1996) *Coronaviruses. Fields Virology*, DMK, Fields BN, Howley PM, *et al.* (eds), Lippincott-Raven **1**: 1095–1103.
7. Snijider EJ, Bredenbeek PJ, *et al.* (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* **331**(5): 991–1004.
8. Kapikian AZ, Estes MK, *et al.* (1996) Norwalk group of viruses. In: DMK, Fields BN, Howley PM, *et al.* (eds), *Fields Virology*. Lippincott-Raven, Vol. 1, pp. 783–810.
9. Theil HJ, Konig M. (1999) Caliciviruses: an overview. *Vet Microbiol* **69**(1–2): 55–62.
10. Prasad BV, Hardy ME, *et al.* (1999) X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**(5438): 287–290.
11. Hollinger FB, Emerson SU. (1996) Hepatitis A. In: DMK, Fields BN, Howley PM, *et al.* (eds), *Fields Virology*. Lippincott-Raven, Vol. 1, pp. 799–840.
12. Purcell R. (1996) *Hepatitis E virus. Fields Virology*. DMK, Fields BN, Howley PM, *et al.* Lippincott-Raven **2**: 2831–2843.
13. Li SW, Zhang J, *et al.* (2005) Mutational analysis of essential interactions involved in the assembly of hepatitis E virus capsid. *J Biol Chem* **280**(5): 3400–3406.
14. Mitsui T, Tsukamoto Y, *et al.* (2004) Prevalence of hepatitis E virus infection among hemodialysis patients in Japan evidence for infection with a genotype 3 HEV by blood transfusion. *J Med Virol* **74**(4): 563–572.
15. Takahashi K, Kitajima N, *et al.* (2004) Complete or near-complete nucleotide sequence of hepatitis E virus genome recovered from a wild boar, a deer, and four patients who ate the deer. *Virology* **330**(2): 501–505.
16. Tei S, Kitajima N, *et al.* (2004) Consumption of uncooked deer meat as a risk factor for hepatitis E virus infection: an age- and sex-matched case-control study. *J Med Virol* **74**(1): 67–70.
17. de Mattos CA, de Mattos CC, Rupprecht CE. (1996) Rhabdoviruses. In: DMK, Fields BN, Howley PM, *et al.* (eds), *Fields Virology*. Lippincott-Raven, Vol. 1, pp. 1245–1277.
18. Hankins DG, Rosekrans JA (2004) Overview, prevention, and treatment of rabies. *Mayo Clin Proc* **79**(5): 671–676.

This page intentionally left blank

Chapter 11

Virus Architecture Probed by Atomic Force Microscopy

A. J. Malkin[†,*], *Yu. G. Kuznetsov*[‡], *M. Plomp*[†] *and A. McPherson*[‡]

Atomic force microscopy (AFM) has recently emerged as an effective complement to X-ray crystallography and electron microscopy techniques for studying virus structure, assembly and function. AFM allows direct, high-resolution visualization of the native surface of both polymorphic and pleiomorphic viruses and their subviral structures. These AFM data, as we demonstrated recently for vaccinia virus, allow successful modeling of the complex architecture of a large virus. AFM can be used to elucidate dynamic processes associated with the lifecycle of viruses *in vitro* and probe the mechanisms of virus entry and budding on the surfaces of virus-infected cells.

Introduction

Currently, X-ray crystallography and electron microscopy (EM) are the two primary tools for structural studies of viruses, and they have provided detailed structures for a wide range of icosahedral and helical

*Corresponding author. E-mail: malkin1@llnl.gov
[†]Chemistry, Materials, and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA.
[‡]Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697-3900, USA.

viruses. In particular, X-ray crystallography[1] is the only available technique allowing structure determination at the atomic level, while recent advances in cryo-EM coupled with developments in image recognition techniques currently allow modeling of virus structures up to ~5 Å resolution.[2] Both techniques provide direct three-dimensional structural information, and they allow visualization of the interior of the virus as well as its surface. Similarly, transmission electron microscopy is currently almost the only technique to probe virus-cell interactions. However, X-ray crystallography and EM have several distinct limitations. X-ray diffraction requires the virus to be in a crystalline form, which is often impossible to achieve. Furthermore, it is currently limited to the studies of viruses with sizes of up to ~ 50 nm. It is unlikely that future technological developments will push this limit beyond 150 nm. Cryo-EM is most useful for highly symmetrical viruses, such as icosahedral viruses. Both techniques require highly purified, structurally homogeneous virus samples. Hence, because of structural heterogeneity, lack of symmetry and large sizes, many animal and human viruses are not amendable to either X-ray diffraction or cryo-EM techniques. Another limitation is that both techniques provide a time-and-space averaged model of the specimen population and therefore cannot visualize unique features that differ from virion to virion. Finally, both X-ray diffraction and cryo-EM require substantial amounts of time for sample preparation, data collection and calculations.

In the past five to seven years, significant advances have been made in structural studies of viruses and virus-infected cells by atomic force microscopy (AFM). The principles of AFM operations have been described in great detail elsewhere (Dufrene, 2004; Malkin *et al.*, 2005).[3,4] AFM has different virtues that tend to be complimentary to those of X-ray diffraction and EM. AFM allows rapid visualization of high-resolution architectures of single virions present in a relatively crude preparation. This enables access to variations in the particle structure within a population, which could be important for the evaluation of their infectivity. AFM provides a resolution comparable to conventional EM methods,[5] while at the same time it can be conducted in

fluid under native conditions. Furthermore, environmental conditions can be easily changed during the course of AFM experiments, allowing study of the structural dynamics of viruses in response to environmental and chemical stimuli.

In this chapter, we will demonstrate the capabilities of AFM in probing the architecture of plant, animal and human viruses representing a wide size range (60 Å – 0.5 μm), imaged under a broad range of conditions (air-dried, fully hydrated, various inorganic and organic solutions) and constellations (single, in a crystal, dissected, budding from cells).

AFM Investigations of High-Resolution Structures of Viruses

The overall morphology of viruses, having a wide range of geometrical structures and spanning the size range from 16 nm for small T = 1 icosahedral satellite viruses[6] to almost half a micron for poxviruses,[7] can be imaged by AFM under physiological conditions (Fig. 1). Furthermore, AFM provides a resolution sufficient not only to visualize the gross shape of virions, but also to unravel the molecular details of their surfaces. Here, we will describe several applications of AFM to probe high-resolution structures and assemblies of a wide range of viruses.

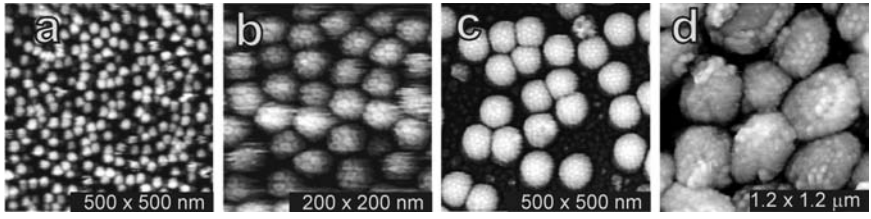


Fig. 1. (**a**) 16 nm diameter satellite tobacco mosaic virus (STMV) particles adsorbed on mica. (**b**) Brome mosaic virus (BMV) particles adsorbed on mica. Capsomeres are evident on the virus surface. (**c**) Ty3 RT mutant virus-like particles (VLPs). (**d**) Intracellular mature vaccinia (IMV) virus particles adsorbed on mica.

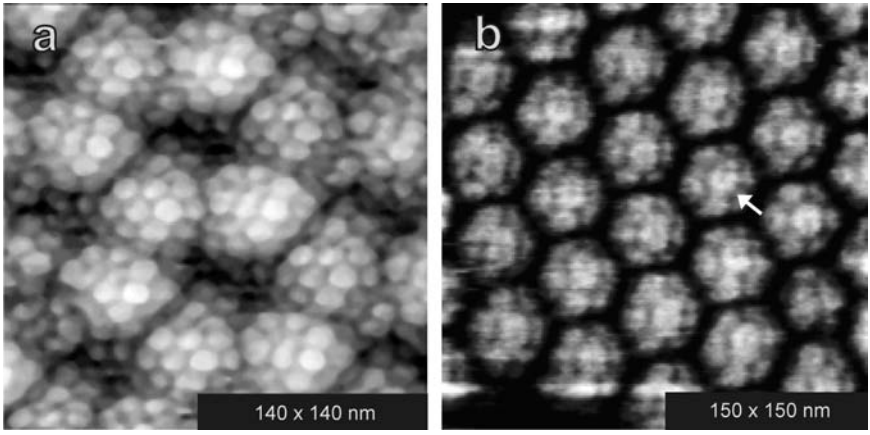


Fig. 2. (**a**) Turnip yellow mosaic virus (TYMV) particles embedded in the surface of a virus crystal exhibit an icosahedral array of capsomeres. (**b**) Brome mosaic virus (BMV) crystal. Individual virions show capsomeres with a depression in their center (*arrow*).

T = 3 Icosahedral Viruses

In work on turnip yellow mosaic virus (TYMV), illustrated in Fig. 2a, it was demonstrated for the first time that AFM allowed the structure of a virion to be visualized to such high resolution that individual protein capsomeres on the virion surface could be clearly resolved.[8]

The capsid of the T = 3 icosahedral TYMV virion is 28 nm in diameter and is composed of 180 identical protein subunits, organized into 12 pentameric and 20 hexameric capsomeres.[9] In AFM images, pentameric and hexameric clusters, which are both roughly 60 Å across, are clearly discriminated. The difference between the highest and lowest points on the capsid surface, about 45 Å,[9] was accurately reflected by AFM measurements. Furthermore, it was demonstrated that viruses of different, but closely related, virus families could be discriminated from each other by AFM on the basis of differences in capsid structure.[6] Thus, while both are 28 nm diameter T = 3 icosahedral viruses, comparison of the capsomere structures of TYMV (*Tymovirus* family) and BMV (brome mosaic virus,

Bromovirus family, Fig. 2b) shows them to be strikingly different. In BMV the capsomeres are wide and have crown-like structures, while those in TYMV are close and dense. Further inspection indicates ~25 Å holes at the centers of BMV capsomeres (Fig. 2b) consistent with the structure of BMV based on X-ray diffraction analysis.[10] This demonstrates that imaging at a lateral resolution of ~20 Å is possible on viruses, *in vitro*.

Herpes Simplex Virus-1

The clarity with which structural detail can be recorded on the surfaces of small T = 3 plant viruses suggested that AFM may be even more broadly useful as an analytical tool for macromolecular structural investigations and may provide important topographical information on large animal and human viruses. Thus the capabilities of AFM in structural studies of human viruses may be illustrated by our work[5] on the visualization of the icosahedral capsid of herpes simplex virus-1 (HSV-1). The structure of the icosahedral capsid of HSV-1, which was deduced by cryo-EM to 8.5 Å,[11] provides a standard against which the AFM images can be compared. The main components of the 125 nm T = 16 icosahedral HSV-1 virion are 12 pentameric capsomeres and 150 hexameric capsomeres. These capsomeres are interconnected by 320 smaller protein complexes known as triplexes.[11]

As illustrated in Fig. 3a, the capsomere structure of the HSV-1 capsid could be visualized by AFM even at relatively low magnification, with the capsomere size of ~15 nm being accurately reflected. At higher resolution, the protein clusters that link adjacent capsomeres and the substructure of individual capsomeres could be further visualized (Fig. 3b). The 4–5 nm clusters between capsomeres correspond to triplexes that are a characteristic feature of the capsid. Furthermore, many capsomeres show a ~35 Å hole in their center, which is consistent with the cryo-EM reconstruction model of the HSV-1 capsid. Furthermore, while the overall structure of capsids seen in AFM images resembles the cryo-EM reconstruction model, natural variations of the mean structure were observed as well. AFM images of

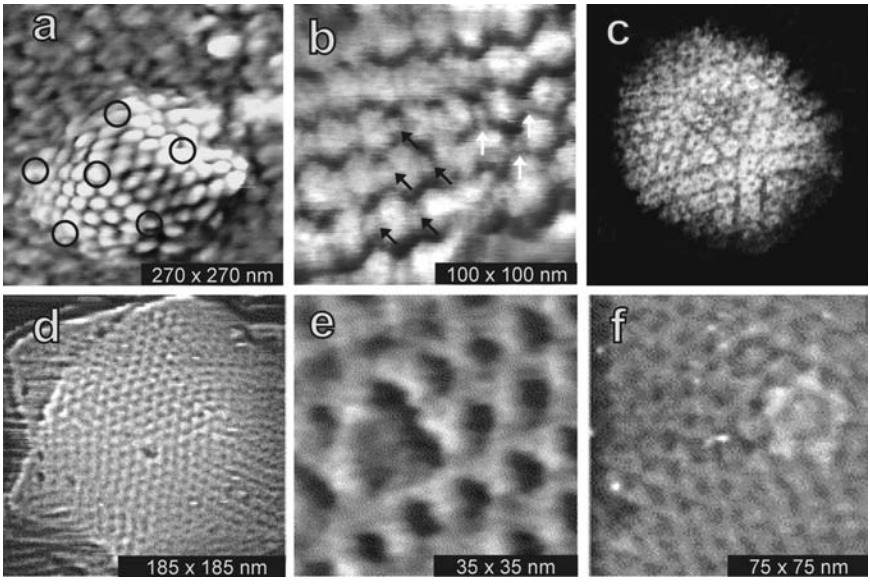


Fig. 3. (**a–c**) Herpes Simplex Virus-1 (HSV-1). (a) HSV-1 capsid with hexons and pentons (circles) capsomeres constituting the icosahedral symmetry. (**b**) High-resolution image of an HSV-1 capsid showing small protein complexes known as triplexes between three (white arrows) as well as two (black arrows) capsomeres. EM image reconstruction demonstrated these units only present in the former location. (**c**) HSV-1 capsid as seen by EM. (Reprinted from http://www.uct.ac.za/depts/mmi/stannard/linda.html. Copyright Linda M Stannard, 1995) (**d–e**) *Paramecium bursaria* chlorella virus type 1 (PBCV-1). (**d**) PBCV-1 capsids display a honeycomb appearance. (**e**) At high resolution, the triangular nature of the trimeric coat protein is seen. Note the hole in the center formed by the loss of one trimer. (**f**) The pentameric arrangement of proteins about the five-fold vertices, with a unique and different protein seen exactly on the vertex.

HSV-1 capsids demonstrate that the AFM technique provides resolution comparable to conventional EM techniques (Fig. 3c). It must be emphasized that the HSV-1 virions were not highly purified; nevertheless, effective analysis of virion components was possible. This demonstrates that AFM visualization can be applied effectively to relatively crude virus preparations.

Paramecium Bursaria Chlorella Virus Type 1

High-resolution AFM imaging of *Paramecium bursaria* chlorella virus type 1 (PBCV-1) allowed visualization of the virion's surface lattice, as illustrated in Fig. 3d.[12] The icosahedral PBCV-1 can be deconstructed into 20 triangular and 12 pentagonal plates known as trisymmetrons and pentasymmetrons,[13] which are centered around threefold and fivefold symmetry axes, respectively. Higher-resolution images, as seen in Fig. 3e, reveal the individual trimeric capsid proteins that form a honeycomb structure, as well as the pentagonal vertices (Fig. 3f). These vertices clearly exhibit a single apical globular protein in their centers and five other unique globular proteins that together form an exact pentamer. These structures are consistent with the model of the PBCV-1 virion as derived from cryo-EM.[13] In addition, AFM studies provide detailed structural information on both the arrangements of trimeric capsid proteins and the pentagonal vertices. Thus, from AFM images, the small hole in the trimer center has a distinctive triangular shape that is more accentuated and angular than the "doughnut" shape deduced from cryo-EM.[14]

In the cryo-EM model a hole was assigned to the pentagonal apex, while AFM images clearly demonstrate the presence of a unique protein in its center (Fig. 4a). The arrangement of a trisymmetron lattice appears not to be affected by regular AFM scanning force. However, purposely higher AFM tip pressure was found to press the apical protein into the vertex, leaving in its place a distinct hole (Fig. 4b). The central protein reappears in the vertex upon decrease of the tip pressure to its original value. This demonstrates the apparent lesser integration of the apex protein in the lattice structure, resulting in the ability to move freely perpendicular to the virion surface.

Ty3 Retrotransposon Virus-like Particles

AFM was successfully utilized for the first determination of the molecular structure of the Ty3 retrotransposon, virus-like particles (VLPs) from *Saccharomyces cerevisiae*.[15] Transmission EM[16] provided the information on the size range of Ty3; however, no detailed

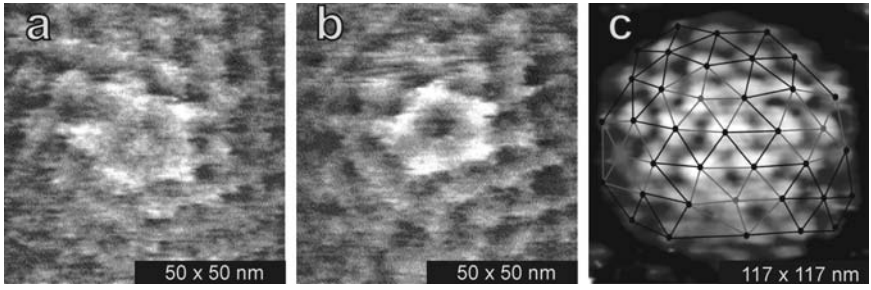


Fig. 4. (**a**) Under normal AFM scanning conditions, the center protein on the PBCV-1 is visible within the 5-protein pentagon. (**b**) With increased tip pressure, the center protein disappears, while the surrounding pentagon stays in place. Upon lowering the tip pressure, the center protein reappears (not shown). (**c**) Icosahedral structure of a wild type Ty3 VLP with icosahedral overlay; pentagonal vertices are shown in grey and hexagonal vertices are shown in black.

structural information had been reported. Wild-type Ty3 particles range in diameter from 25 nm to 52 nm. This particle variability precludes successful structural analyses by cryo-EM, which relies on molecular averaging. AFM imaging of Ty3 VLPs demonstrates the spherical shape of a vast majority of the Ty3 particles, as well as capsomere arrangements on their surfaces (Fig. 4c). Measurements of the capsomere center-to-center distances of a large number Ty3 VLPs yields an average distance of 11.4 nm, consistent with a T = 7 icosahedral triangulation number. Additionally, the number of hexameric capsomeres lying between pentameric capsomeres, as illustrated in Fig. 4c, is consistent with T = 7 as well. Since T = 7 icosahedra are non-centrosymmetric they can exist in two enantiomorphs, *d* and *l*. From the height information preserved in the AFM images, further analysis allowed assignment of the Ty3 VLPs to the T = 7*d* class.

Vaccinia Virus

AFM can provide important high-resolution structural information on polymorphous asymmetric viruses, which are not amendable to

high-resolution EM-reconstruction analysis. This was demonstrated in AFM studies of structure and assembly of intracellular mature vaccinia (IMV) virus.[7] Vaccinia virus, a homologue of smallpox and the basis of the smallpox vaccine, is one of the largest viruses to replicate in humans. Among the human-tropic viruses, vaccinia has one of the most structurally complex, markedly asymmetric virions. The structure of the vaccinia virion has been extensively studied for several decades by conventional EM.[17–19] The virion was initially described as brick-shaped with dimensions of approximately $300 \times 230 \times 120$ nm, and with inner and outer membranes sandwiching a pair of lateral bodies and enclosing a central core containing the genomic DNA.[17] Conventional EM studies revealed membranous surface tubules and a dumbbell shape of the core. These structural attributes were not visible in cryo-EM images,[20] which suggested that these features may have been artifacts of virion dehydration and non-isotropic collapse during conventional EM sample preparation. Thus, because of these inconsistencies, the combination of conventional and cryo-EM did not provide a definitive model for the structure of the vaccinia virion.

AFM images of fully hydrated vaccinia virions (Fig. 1d) demonstrate that the majority of virions have an ellipsoidal shape with major and minor axial dimensions in the range 320–380 nm and 260–340 nm, respectively, and a virion height of 240–290 nm. While the lateral dimensions derived from AFM images are consistent with those from cryo-EM studies, the AFM virion height is approximately double that of the cryo-SEM- and EM-estimated heights of 110 nm and 150 nm, respectively (Griffiths *et al.*, 1994).[21] Note that while height information from AFM is very precise, with a resolution approaching 0.1 nm, estimation of vertical dimensions by EM techniques is not straightforward and depends on the stage tilt angle. Fully hydrated IMV particles, as revealed by AFM, have a "rounded barrel" shape, which differs from the classical "brick-shaped" morphologies for pox viruses described in the literature.[17]

It is expected that due to its thickness and electron-translucent properties, visualization of the high-resolution surface morphology of the vaccinia virion by cryo-EM becomes difficult, since EM images provide a projection of the entire thickness of the sample. Indeed,

high-resolution cryo-EM images,[20] and even thin section EM images[22] provide a rather smooth, featureless topographical appearance for the virion. In contrast, high-resolution *in situ* AFM images demonstrate a high density of protrusions on the virion surface, giving it a "knobby" appearance. The protrusions appear to be of a fairly uniform size of about 25 × 30 nm, though irregular in their arrangement. In some of the most detailed images (Fig. 5a), subunits of ~6 nm diameter can be discriminated within the protrusions. This would correspond to a molecular weight of about 120–150 kDa for a monomeric globular protein. Since the molecular weights of the 16 major membrane

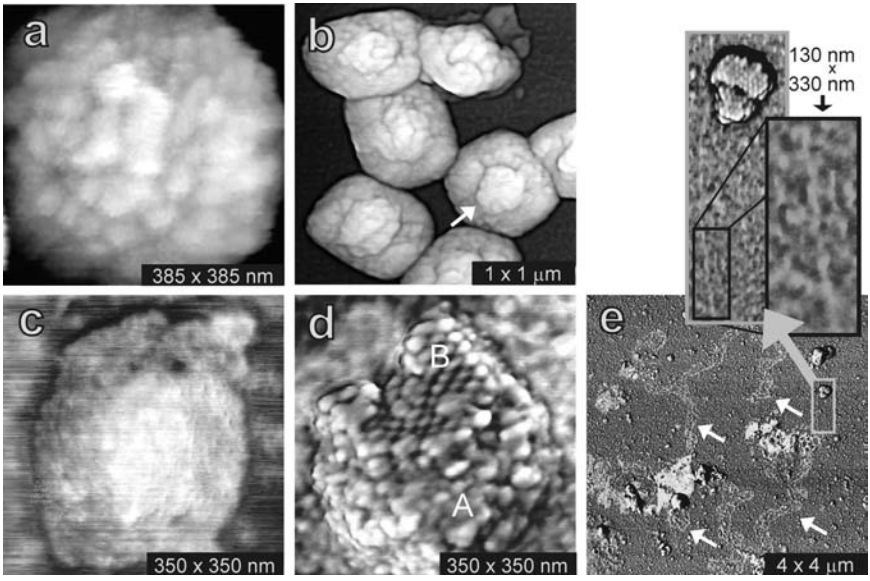


Fig. 5. (**a**) High-resolution image of fully hydrated IMV virion in buffer. (**b**) Air-dried IMV virions show a central raised area (arrow). (**c**) HSV-1 virion still covered in its lipid envelope. (**d**) Addition of 0.2% Triton X-100 results in removal of the envelope, showing tegument proteins (A) and the capsid with individual capsomeres (B). (**e**) DNA escapes from HSV-1 virions after treatment with 0.5% SDS (white arrows). The area in the grey rectangle is shown in the two insets at higher resolution. Here, an individual double-stranded DNA string is seen to escape a damaged virion.

proteins of vaccinia virus are all considerably less,[23] the subunits probably represent oligomers.

AFM allows direct visualization of the structural dynamics of viruses in response to changes in their environment. In this particular case, AFM allows imaging of IMV particles in both water and air. Thus, as illustrated in Fig. 5b, upon dehydration air-dried vaccinia virions assumed more rectilinear shapes with relatively uniform dimensions of 300–360 × 240–280 nm (lateral) × 120–130 nm (height), representing a reversible 2.2–2.5-fold shrinkage. This shrinkage is accompanied by topological and topographical changes resulting in "brick-shaped" morphologies similar to those observed in EM images. The dried virions have a pronounced, raised area in their center, which extends approximately 30 nm above the virion surface. Thus, AFM observations directly demonstrate that conventional EM images shown in literature that purported to show native vaccinia virion structure were fraught with drying artifacts. AFM also demonstrates the remarkably deformable, fluid nature of the virion structure, which allows a ~2.5-fold collapse along the vertical dimension upon drying.

Visualization of Subviral Structures

In the previous sections we described applications of AFM for probing the high-resolution structures of the virions' outer surfaces. In addition, AFM can be utilized to probe their internal structural features. This can be accomplished through a controlled dissection of viruses with chemical agents or enzymes, resulting in the sequential peeling away of layers of structure. Newly revealed/exposed structural layers can then be imaged by AFM.

Herpes Virus

The ability to utilize AFM for imaging of subviral structures was first demonstrated in the visualization of herpes simplex virus-1.[5] The outer layer of the HSV-1 virus is formed by an oversized lipid envelope (Fig. 5c). Inside is the actual capsid, as well as the tegument, a varied collection of proteins present between envelope and capsid.

Mixing of a virus solution with detergent (0.2% Triton X-100) results in the removal of the lipid envelope from intact virions, which then renders the underlying tegument and capsomere structure visible (Fig. 5d). On the upper part of the capsid the underlying, highly regular capsomere packing can be recognized, while its lower part is covered with an irregular coating of particles of up to 10 nm. This ~10 nm size corresponds well with an MW of 336 kDa of the largest protein VP 1–3 present in the tegument. More vigorous treatment of virions with detergent results in the loss of all envelope and tegument, revealing the complete icosahedral capsid structure as seen in Figs. 3a and 3b.

AFM can probe not only the structures of large macromolecular assemblies, but also their dynamic processes. For example, the disassembly of virions and the release of HSV DNA were observed as illustrated in Fig. 5e. Here, the treatment of HSV-1 virions with 0.5% sodium dodecyl sulfate resulted in the emergence of DNA in the form of bundles of many intertwined DNA chains. Individual DNA fibers were found to have diameters of 1.5–3 nm, which corresponds well to the 2.5 nm diameter of a double-stranded DNA double helix. In the inset of Fig. 5e entire capsomeres are seen to be lost from the capsid surface, and an individual double-stranded DNA chain emerges from the capsid.

Vaccinia Virus

In the case of vaccinia virus the internal virion structures were revealed through the subsequent treatment of virions with sets of chemicals and/or enzymes.[7] Consecutive subviral structures including core envelopes, viral cores, and nucleoprotein filaments, as well as viral DNA, were visualized (Fig. 6). Note that the "studs" seen in Fig. 6a on the enveloping membrane surface have the same size as the protrusions on the surfaces of untreated IMV (Fig. 1d). Membrane patches varied in thickness from 25 nm to 35 nm and were ~700 × 900 nm in surface area.

The heights of the viral cores seen in Fig. 6b are in the range of 170–220 nm, which is roughly the height of the intact IMV virion

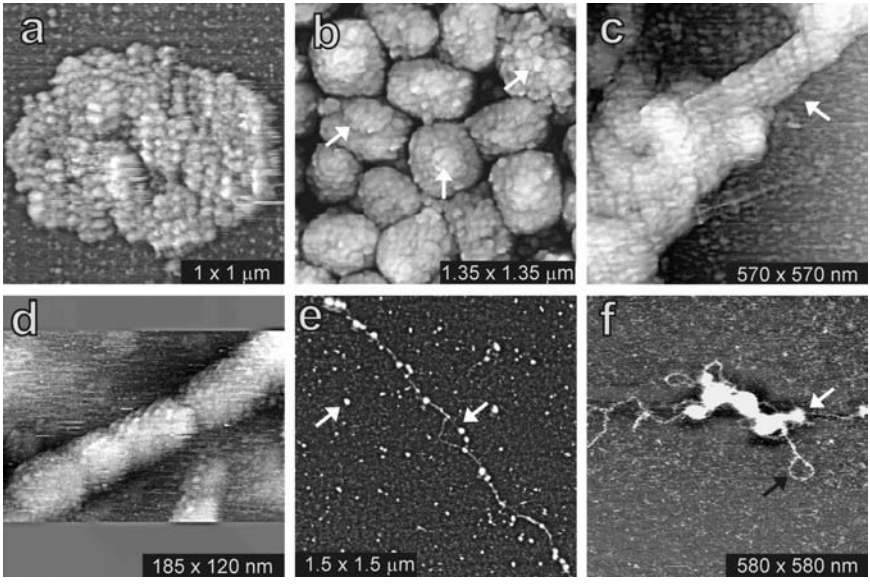


Fig. 6. Dissection of IMV. (**a, b**) Treatment of non-ionic detergent (1% Igepal) in combination with a reducing agent (2% 2-mercaptoethanol) for 30–45 min at 37°C leads to separation of core envelopes (**a**) and virion cores (**b**). (**c**) Extended treatment (120–180 min) with the same combination leads to the unfolding of the virion cores, leading to 30–40 nm-diameter tubules (arrow). (**d, e**) Further dissection using proteinase K at 37°C for 90–180 min results in the appearance of the 16 nm-diameter segmented tubules (**d**) and residual tubular segments, either isolated or associated with DNA strands (white arrows in (**e, f**)).

lacking a 30 nm thick envelope. As seen in Fig. 6b, many viral cores possess satellite domains. These satellite domains may correspond to the lateral bodies observed in conventional EM images.[17]

Prolonged treatment of IMV viruses with non-ionic detergent plus reducing agent produces complete unfolding of IMV and the emergence of 30–40 nm diameter tubules as seen in Fig. 6c. On these tubules, a helical array of protein subunits with left-handed helicity and a pitch of approximately 16 nm is evident. Treatment with high proteinase K concentrations plus SDS results in the same 30–40 nm

tubular structures. However, in this case, disintegration continues beyond this stage, leading to the exposure and release of the components within the tubules.

The components initially revealed are filaments of 16 nm in diameter (Fig. 6d). At higher resolution, the surfaces of the 16 nm filaments, like those of the 30–40 nm tubules, exhibited a helical geometry. In addition, the filaments were linearly segmented (Fig. 6d).

With high proteinase K concentrations and longer exposure, a further disintegration of the virus was promoted. What remained were long strands of double helical DNA, associated with residual portions of the 16 nm-diameter filaments (Fig. 6e). As seen in Fig. 6f, the DNA strands were observed emerging from the segments of the 16 nm tubules, indicating that the DNA is packed within the 16 nm filaments.

AFM observations from chemical and enzymatic dissection lead to a novel structural model of the vaccinia virion based on a hierarchy of observed substructures.[7] The double-stranded genomic DNA is encapsidated within a segmented protein sheath, forming an extended "filament" of ~16 nm diameter with a helical surface topography (Fig. 6d). This filament is in turn enclosed within a 30–40 nm diameter "tubule," which also shows a helical topography (Fig. 6c). The ellipsoidal virion core apparently arises from a folded, condensed arrangement of the 30–40 nm tubules (Fig. 6b) surrounded by a shell heavily studded with proteins. Proteins visualized attached to the tubules may mediate folding/compacting of the latter and/or represent vestiges of the core shell. The condensed tubular mass combines with a 50–70 nm-diameter domain, and this pairing of core and satellite domain is then enshrouded by membranes heavily studded by proteins and containing protrusions and spicules on their surface.

AFM Visualization of Virus-infected Cells

AFM provides an important capability to study mechanisms of cell infection with viruses. These mechanisms include the membrane fusion processes that lead to entry and exit of viruses into and out of the cell, as well as morphological changes of the cell surface

associated with these events. Currently, *in situ* high-resolution AFM imaging of virus-infected cells is rather challenging due to the inherited softness of cell membranes, and their subsequent deformation. However, these problems were successfully overcome using standard histological procedures (fixing the cells with glutaraldehyde and osmium tetroxide), which resulted in the preservation of virus and cell membrane structures.[24]

Mouse Leukemic Virus-infected Cells

In vitro AFM analysis of mouse leukemic virus (MuLV) infected NIH 3T3 cells[24] allowed visualization of the emergence of viruses in all stages of the budding process. The distribution of sizes of emerging virions was found to be broad, with an average particle diameter of 145 nm. The surfaces of the virions (Figs. 7a and 7b) were found to be studded with densely packed "tufts" of protein, having sizes of 11–12 nm.[25] The number of "tufts" on the virion surfaces was estimated to be ~100 ± 20. AFM-based immuno-labeling experiments[25] demonstrated binding of monoclonal antibodies directed against the SU domain of the envelope protein to these "tufts" on the virion surface. This indicates that the "tufts" are formed by aggregates of the SU protein. The absence of a regular arrangement of the "tufts" on the virion surface as revealed by AFM is consistent with the cryo-EM observed pleiomorhic appearance of MuLV virions.[26] In particular, the absence of trimeric symmetry in the tufts is rather intriguing, since it is commonly assumed that the SU protein is organized into threefold symmetrical trimers.[27]

AFM studies of M-MuLV virions lacking the gene for the envelope glycoprotein (env^-) demonstrated that the surface architectures of env^- particles (Fig. 7c) is profoundly different compared with that for wild-type virions.[25] In particular, the protein "tufts," which appear to be a structural attribute of wild-type virions, are not seen on the surface of env^- MuLV virions. Furthermore, it was found that the surfaces of env^- virions exhibit less robust mechanical properties compared to the surfaces of wild-type virions. Thus, the surface architecture of env^- virions was found to be strongly dependent

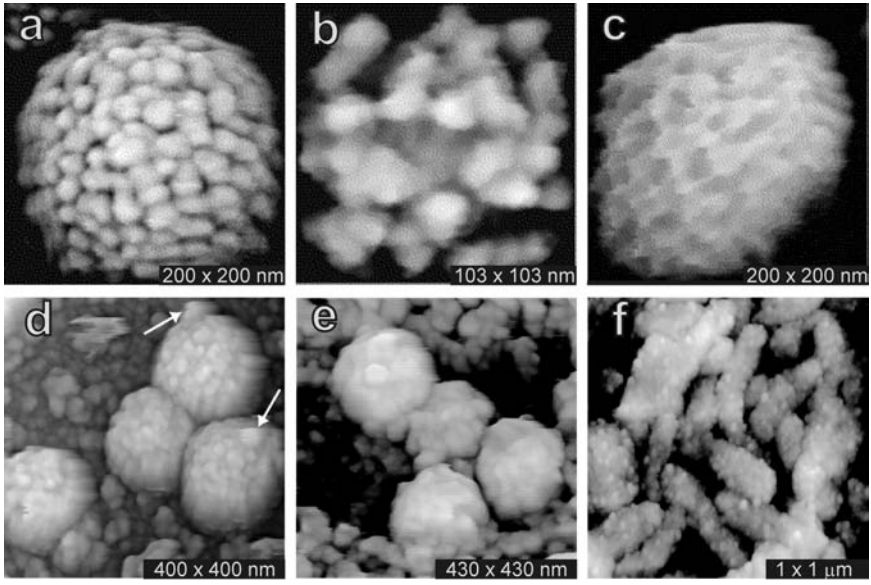


Fig. 7. Moloney murine leukemia virus (M-MuLV). (**a, b**) Wild-type env^+ virion exhibiting dense arrangements of protein "tufts" on its surface: (**a**) whole viron and (**b**) high-resolution. (**c**) A virion of the M-MuLV env^- mutant lacking the gene for the envelope glycoprotein shows indeed a "bald" surface lacking the surface proteins. The irregular surface character is a consequence of the deformable and unstable membrane surface. (**d**) Isolated MuLV virions show the appearance of protruding blebs (arrows), never more than one per virion, that are likely to be scars resulting from the budding process. (**e**) MuLV virions emerging from an infected NIH 3T3 fibroblast cell. The granular background is the 3T3 cell surface. (**f**) The surface of an NIH 3T3 cell infected with the MuLV mutant $gPr80^{gag}$. The emerging mutant virions are unable to mature and separate from the cell surface, and instead form prominent tube-like protrusions on the cell surface. The cells in (**e, f**) were fixed with gluteraldehyde and dehydrated with, and imaged under, ethanol to enable AFM imaging.

upon the force exhibited by the AFM probe during scanning. It was suggested that the surfaces of the env^- virions contain a lower density of associated proteins, which have markedly unstable mechanical properties.

The surface architecture of isolated MuLV virions (Fig. 7d) appears to be similar to that seen on the virions budding from the surfaces of cells (Fig. 7e).[28] A common structural feature of isolated MuLV virions was a knob-like protrusion, which more likely represents the "bud scar" resulting from the pinching off of the cell membrane. Partially damaged particles lacking portions of viral envelope also allowed visualization of intact viral cores with diameters of 65 nm, and enabled an estimate of the viral shell thickness of 35–40 nm.

NIH 3T3 cells infected with wild-type MuLV virus exhibit a number of budding viruses on their surfaces (Fig. 7e). In contrast, cells infected with virus carrying the gPr80gag mutations were found to be covered with tubular structures (Fig. 7f).[24] The diameter of tubular structures was found to be comparable to ones measured for wild-type MuLV virions. Similarly, the surface features of the tubular gPr80gag mutation structures appear to be similar to ones seen on the surfaces of wild-type virions. It was concluded that the observed tubular structures on the cell surfaces are a result of the defective morphogenesis of the mutant virions and their subsequent failure in budding from the cell's surface.

Human Immunodeficiency Virus-infected Cells

The first AFM images of isolated human immunodeficiency virus (HIV) and HIV-infected lymphocytes in culture have been recently reported by Kuznetsov *et al.*[29] (Fig. 8a). It was demonstrated that HIV virions are roughly spherical and have a pleomorphic exterior with a dense packing of protruding protein units. The average diameter of virions was found to be about 120 ± 25 nm, which is ~ 10% smaller than that determined from cryo-EM.[30] Similar particles (Figs. 8b and 8c) having diameters of 127 ± 30 nm were observed on infected cell surfaces. Morphologies of these particles varied more than they do for free virions, which is presumably so because the cell surface virions are in various stages of the budding process.

The ~ 20 nm tufts visualized on the surface of HIV particles represent aggregates of the receptor-binding protein gp120, which is commonly assumed to be a symmetrical trimer.[31] However, variations in size

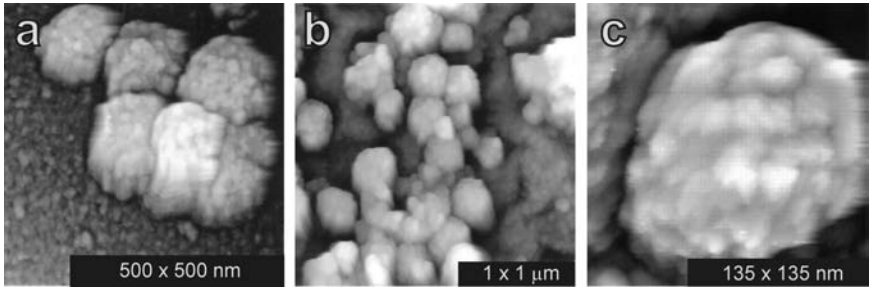


Fig. 8. (**a**) Human immunodeficiency virus (HIV) particles isolated from culture medium. (**b**) The surface of an infected H9 lymphocytic cell is covered with a cascade of HIV particles. (**c**) High-resolution AFM image of an HSV virion on an H9 cell surface. Clusters of envelope proteins seem to arise from more or less arbitrary associations of subunits. Average particle diameter is 127 nm.

and shape of tufts along with the absence of threefold symmetry as seen in AFM images suggest that in the native HIV particle, gp120 may form arbitrary clusters rather than closely associated symmetric units.

Acknowledgments

The authors thank E. K. Wagner, P. D. Gershon, H. Fan, W. E. Robinson, S. Sandmeyer and J. V. Van Etten for virus preparations and helpful discussions. We wish to thank A. Greenwood, R. Lucas, J. Zhou, M. K. Rice for technical assistance. The National Aeronautics and Space Administration and National Institute of Health supported this research. Work was performed under the auspices of the US Department of Energy by the University of California, Lawrence Livermore National Laboratory, under contract W-7405-ENG-48.

References

1. Harrison SC. (1990) Principles of virus structure. In B.N. Fields and D.M. Knipe (eds), *Fields Virology*, 2nd edn., pp. 37–61. Plenum Press, New York.

2. Baker TS, Olson NH, Fuller SD. (1999) Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbial Mol Biol Rev* **63**: 862–922.
3. Dufrene YF. (2004) Using nanotechniques to explore microbial surfaces. *Nature Rev Microbiol* **2**: 451–460.
4. Malkin AJ, Plomp M, McPherson A. (2005) Unraveling the architecture of viruses by high-resolution atomic force microscopy. In P.M. Liberman (ed), *DNA viruses. Methods and Protocols*, pp. 88–108. Humana Press, Totowa.
5. Plomp M, Rice MK, Wagner EK, *et al.* (2002) Rapid visualization at high resolution of pathogens by atomic force microscopy: structural studies of herpes simplex virus-1. *Am J Pathology* **160**: 1959–1966.
6. Kuznetsov YuG, Malkin AJ, Lucas RW, *et al.* (2001) Imaging of viruses by atomic force microscopy. *J Gen Virol* **82**: 2025–2034.
7. Malkin AJ, McPherson A, Gershon PD. (2003) Structure of intracellular mature vaccinia virus visualized by *in situ* atomic force microscopy. *J Virol* **77**: 332–340.
8. Malkin AJ, Kuznetsov YuG, Lucas RW, McPherson A. (1999) Surface processes in the crystallization of turnip yellow mosaic virus visualized by atomic force microscopy. *J Struct Biol* **127**: 35–43.
9. Canady MA, Larson SB, Day J, McPherson A. (1996) Crystal structure of turnip yellow mosaic virus. *Nature Struct Biol* **3**: 71–781.
10. Lucas RW, Kuznetsov YuG, Larson SB, McPherson A. (2001) Crystallization of brome mosaic virus and T = 1 brome mosaic virus particles following a structural transition. *Virology* **286**: 290–303.
11. Zhou ZH, Dougherty M, Jakana J, *et al.* (2000) Seeing the herpesvirus capsid at 8.5 Å. *Science* **288**: 877–880.
12. Kuznetsov YuG, Gurnon JR, Van Etten JL, McPherson A. (2005) Atomic force microscopy investigation of a chlorella virus, PBCV-1. *J Struct Biol* **149**: 256–263.
13. Simpson AA, Nandhagopal N, Van Etten JL, Rossmann MG. (2003) Structural analyses of Phycodnaviridae and Iridoviridae. *Acta Crystallogr D Biol Crystallogr* **59**: (Pt 12), 2053–2059.
14. Yan X, Olson NH, Van Etten JL, *et al.* (2000) Structure and assembly of large lipid-containing dsDNA viruses. *Nature Struct Biol* **7**: 101–103.

15. Kuznetsov YuG, Zhang M, Menees TM, *et al.* (2005) Investigation by atomic force microscopy of the structure of Ty3 retrotransposon particles. *J Virol* **79**: 8032–8045.
16. Hansen LJ, Chalker DL, Orlinsky KJ, Sandmeyer SB. (1992) Ty3 *GAG3* and *POL3* genes encode the components of intracellular particles. *J Virol* **66**: 1414–1424.
17. Fenner F, Wittek R, Dumbell KR. (1989) *The orthopoxviruses.* Academic Press, Inc., New York.
18. Griffiths G, Roos N, Schleich S, Locker JK. (2001) Structure and assembly of intracellular mature vaccinia virus: thin-section analysis. *J Virol* **75**: 11056–11070.
19. Griffiths G, Wepf R, Wendt R, *et al.* (2001) Structure and assembly of intracellular mature vaccinia virus: isolated-particle analysis. *J Virol* **75**: 11034–11055.
20. Dubochet J, Adrian M, Richter K, *et al.* (1994) Structure of intracellular mature vaccinia virus observed by cryoelectron microscopy. *J Virol* **68**: 1935–1941.
21. Roos N, Cyrklaff M, Cudmore S, *et al.* (1996) A novel immunogold cryoelectron microscopic approach to investigate the structure of the intracellular and extracellular forms of vaccinia virus. *EMBO J* **15**: 2345–2355.
22. Hollinshed M, Vanderplassen A, Smith GL, Vaux DJ. (1999) Vaccinia virus intracellular virions contain only one lipid membrane. *J Virol* **73**: 1503–1517.
23. Jensen ON, Houthaeve T, Shevchenko A, *et al.* (1996) Identification of the major membrane and core proteins of vaccinia virus by two-dimensional electrophoresis. *J Virol* **70**: 7485–7497.
24. Kuznetsov YuG, Datta S, Kothari NH, *et al.* (2002) Atomic force microscopy investigation of fibroblasts infected with wild-type and mutant murine leukemia virus (MuLV). *Biophys J* **83**: 3665–3674.
25. Kuznetsov YuG, Low A, Fan H, McPherson A. (2004) Atomic force microscopy investigations of wild-type Moloney murine leukemia virus particles and virus particles lacking the envelope protein. *Virology* **323**: 189–196.
26. Yeager M, Wilson-Kubalek EM, Weiner SG, *et al.* (1998) Supramolecular organization of immature and mature murine leukemia virus revealed by electron cryo-microscopy: implications for retroviral assembly mechanisms. *Proc Natl Acad Sci USA* **95**: 7299–7304.

27. Poignard P, Saphire EO, Parren PW, Burton DR. (2001) GP120: biologic aspects of structural features. *Annu Rev Biochem* **19**: 253–274.
28. Kuznetsov YuG, Low A, Fan H, McPherson A. (2005) Atomic force microscopy investigation of isolated virions of murine leukemia virus. *J Virol* **79**: 1970–1974.
29. Kuznetsov YuG, Victoria JG, Robinson Jr WE, McPherson A. (2003) Atomic force microscopy investigation of human immunodeficiency virus (HIV) and HIV-infected lymphocytes. *J Virol* **77**: 11896–11909.
30. Briggs JAG, Wilk T, Welker R, *et al.* (2003) Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J* **22**: 1707–1715.
31. Wilk T, Gross I, Gowen BE, *et al.* (2001) Organization of immature human immunodeficiency virus type 1. *J Virol* **75**: 759–771.

This page intentionally left blank

Chapter 12

Filovirus Assembly and Budding

Takeshi Noda[*,‡] *and Yoshihiro Kawaoka*[*,†,‡,§]

Filoviridae are single-stranded, non-segmented, negative-sense RNA viruses. To date, only two members of this virus family have been identified: the Ebola and Marburg viruses. Both viruses possess lipid envelopes and have characteristic filamentous shapes. Filovirus infections are a growing concern because they can cause significant morbidity and mortality (30%–90%) and no vaccines or treatments are currently available. In addition, the potential use of filoviruses as biological weapons has prompted the need to develop countermeasures. Although many questions about filovirus pathogenicity remain unanswered, much has been learned in recent years about the filovirus lifecycle, particularly the roles of the viral and cellular components in assembly and budding. Such data have greatly increased our understanding of filovirus particle formation. This review summarizes our current knowledge of filovirus particle formation and the viral proteins believed to be critical for this process.

Introduction

Filoviruses, namely the Ebola and Marburg viruses, are taxonomically classified in the family *Filoviridae*. All filoviruses possess single-stranded, non-segmented, negative-sense RNA genomes.[1] Their virions are

*International Research Center for Infectious Diseases.
†Division of Virology, Department of Microbiology and Immunology, and Institute of Medical Science, University of Tokyo, Shirokanedai, Minato-ku, Tokyo 108-8639.
‡Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama 332-0012.
§Department of Pathological Sciences, School of Veterinary Medicine, University of Wisconsin–Madison, Madison, Wisconsin 53706.

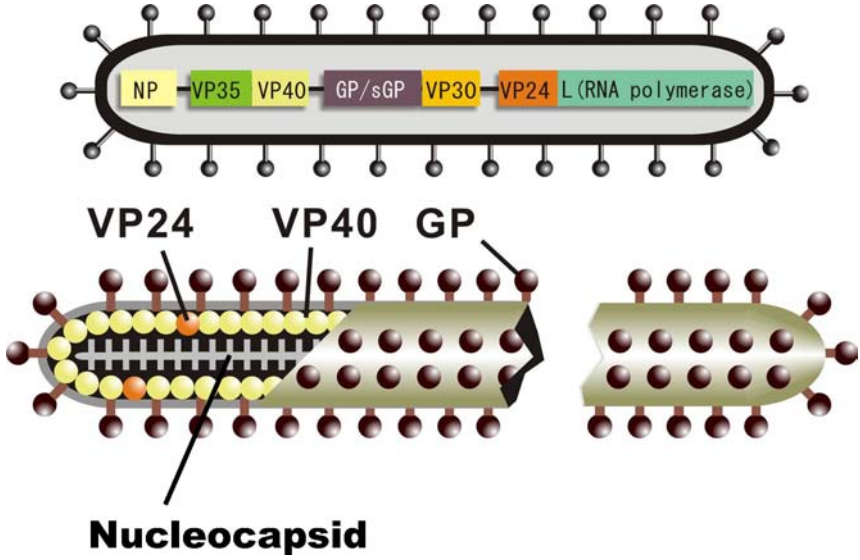


Fig. 1. Schematic diagram of Ebola virus. The Ebola virus genome encodes seven structural proteins and a nonstructural protein (sGP). The virions appear as filamentous particles of approximately 80 nm in diameter and differ greatly in length. The virus particles contain a single surface glycoprotein (GP). The VP40 and possibly the VP24 protein form a matrix underneath the lipid bilayer. The viral nucleocapsid is composed of the viral RNA, NP, VP30, VP35, and L proteins, and may also include the VP24 protein.

characteristically shaped as filamentous particles of approximately 80 nm in diameter with no uniform length[2] (Fig. 1). The 19 kb genome is encapsidated by the nucleocapsid, which consists of the nucleoprotein (NP), viral structural protein (VP) 30 and VP35, and the polymerase protein (L). The nucleocapsid is surrounded by a lipid envelope derived from cellular membranes. Beneath the inner leaflet of the viral envelope lies a lattice of matrix protein (VP40) and possibly VP24, a small membrane-associated protein that may also be a component of the nucleocapsid. Filovirus particles also contain a single, trimeric surface glycoprotein (GP).

Virus assembly is believed to be a step-wise process by which viral components are transported to designated locations for assembly into larger complexes such as nucleocapsids, and ultimately into virus particles. Studies on the assembly mechanisms of viruses of the same

order, *Mononegavirales*, such as paramyxoviruses and rhabdoviruses, have stimulated research on filovirus assembly, and vice versa. Such studies have shown that viruses co-opt cellular transport machinery for their assembly and budding processes, and that even diverse viruses can exploit these pathways.[3]

Virion Budding

Electron microscopic observations suggest that filovirus virions typically bud from the plasma membrane,[2] where viral components such as GP, VP40, and nucleocapsids are assembled. The egress of virions occurs at cytoplasmic vacuoles in monocytes/macrophages,[4] although it remains unclear whether virions that accumulate in intracellular vacuoles are released normally to extracellular spaces. Some electronmicrographs suggest that filamentous virions bud both vertically and horizontally from the plasma membranes of virus–infected cells.[2] Whether differences exist between these two forms and the biological significance of this finding remain unknown.

Matrix proteins play a central role in the virion formation of many enveloped negative-strand RNA viruses.[5] In general, these proteins intrinsically associate with cellular membranes and are believed to form a bridge between the nucleocapsids and the cytoplasmic tails of transmembrane glycoproteins.[5] The Ebola virus and Marburg virus VP40 proteins interact with the plasma membrane and with intracellular membranes.[6–8] Coexpression of VP40 and GP results in the formation of virus-like particles (VLPs) that contain orderly arranged GP spikes,[9] which suggests that VP40 and GP interact. A recent study also found an interaction between VP40 and VP35, a component of the viral nucleocapsid.[10] Thus, VP40 appears to connect the nucleocapsid to the viral shell, similar to the function of the matrix proteins of other negative-strand RNA viruses.

The VP40 Protein

The VP40 matrix protein is the most abundant protein in filovirus particles. The X-ray crystallographic structure of monomeric VP40

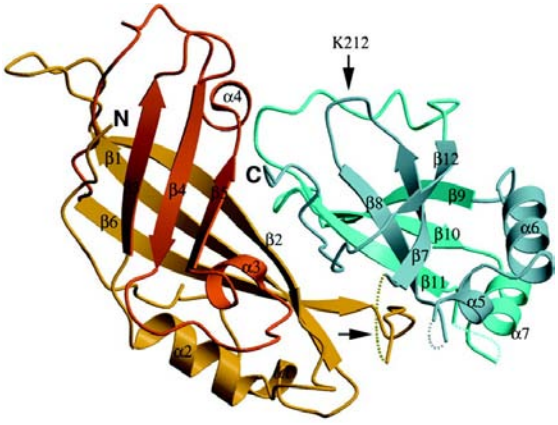


Fig. 2. Ribbon diagram of Ebola virus VP40. Two amino- and carboxy-terminal domains can be seen in red and green, respectively. Courtesy of Dessen *et al.*, 2000.

from Ebola virus has been resolved to 2.0 Å[11] and has revealed that VP40 consists of two structurally homologous domains that comprise six antiparallel β-strands arranged in two β-sheets (Fig. 2). The amino- and carboxy-terminal domains are oriented at an angle of 60°, and are connected by a large flexible loop.[11] VP40 binds strongly to lipid membranes. This binding is resistant to high salt concentrations and appears to be mediated by the C-terminal domain,[6,8] which contains large hydrophobic patches.[11]

The current model predicts that VP40 exists in the cytoplasm as a monomer that is stabilized by interactions between its amino- and carboxy-terminal domains. Destabilization of these interactions liberates the amino-terminal domain to interact with other VP40 molecules and thus form oligomers. This destabilization can be achieved experimentally by a carboxy-terminal truncation of seven amino acids that induces spontaneous hexamerization and the formation of ring-like structures, as demonstrated by chemical cross-linking and electron microscopy,[12] or by membrane association.[12] After the liberation of the amino-terminal domain, VP40 is thought to adopt dimeric and octameric forms; the dimeric VP40 appears to be the building block for both hexameric and octameric structures.[13] The carboxy-terminal

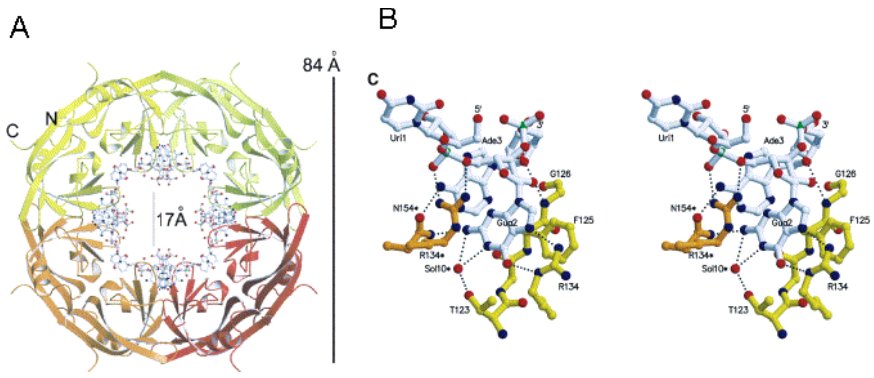


Fig. 3. Octameric ring-like structure adopted by Ebola virus VP40 after its association with specific, single-stranded RNA. (**A**) Ribbon drawing of the ring structure. (**B**) Close-up stereo view of VP40-RNA interactions. Courtesy of Gomis-Ruth *et al.*, 2003.

domain of VP40 may thus prevent VP40 aggregation in the cytoplasm, while VP40 association with the plasma membrane may trigger VP40 oligomerization for the formation of virus particles.

Recent data indicate that VP40 octamerization is critical for Ebola virus replication.[14] Dimeric structures in the VP40 octamer form a pocket at their interface, which binds a short, single-stranded RNA molecule with the sequence: 5′-U-G-A-3′. This RNA-dimer complex helps to stabilize the structure[15] (Fig. 3). Two amino acids, Phe125 and Arg134, are critical for RNA binding. Their replacement impedes octamerization, but it does not affect VLP formation or virion morphology.[14] A recombinant virus containing an Arg134-to-Ala replacement in VP40 could not be recovered,[14] confirming that VP40 octamer formation is likely critical for the Ebola viral lifecycle.

Assembly of VP40 and GP

Several groups have used recombinant DNA technology to demonstrate that Ebola virus VP40 expressed in mammalian cells is released into the culture medium in membrane-bound form.[6,9,13,16] The virosome-like particles containing VP40 were verified by flotation analysis and a

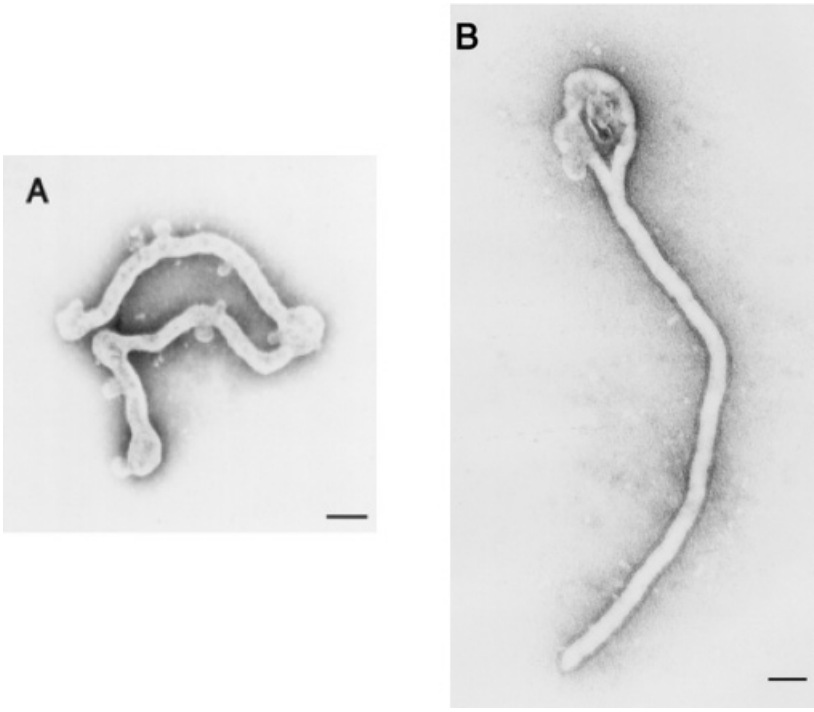


Fig. 4. Filamentous particles comprising VP40 and the lipid bilayer. (**A**) Expression of VP40 in mammalian cells induces filamentous particles similar to spikeless Ebola virions. (**B**) Ebola virus-like particles are formed by coexpression of VP40 and GP.

trypsin-resistance assay. Electron microscopic studies of the virosomes released from VP40-expressing cells also indicated that these particles bore a strong resemblance to GP-minus (spikeless) filovirus particles (Fig. 4a).[9,13] Removal of the carboxy-terminus of Ebola virus VP40 severely compromises the release of virus-like particles,[6] emphasizing the importance of this region in budding. Collectively, these studies have demonstrated that VP40 is the primary morphological determinant in filovirus morphogenesis and that all of the information for membrane association, self-assembly, filamentous particle formation, and budding resides in the sequence of this protein.

For several viruses, their matrix proteins associate with the cytoplasmic tails of the viral envelope glycoproteins.[5] By the same token,

coexpression of GP with VP40 in mammalian cells produces filamentous virus-like particles decorated with GP spikes (Fig. 4b).[9,17,18] These GP spikes appear to be arranged in a regular order.[9] Mature Ebola virus GP is also released in the form of spiked virosomes from cells expressing GP alone;[9,19] however, the arrangement of these GP spikes seems to be irregular.[9] Such morphological findings suggest an interaction between VP40 and GP, although a direct interaction has not yet been demonstrated. We also found efficient incorporation of Marburg virus GP and vesicular stomatitis virus glycoprotein into Ebola VP40-induced VLPs.[20] Since the cytoplasmic tails of these glycoproteins share no sequence similarity, a specific interaction between VP40 and the glycoprotein may not be required for particle incorporation; rather, cellular components, which could exist in lipid raft domains (see below), may link VP40 to the glycoprotein.

Lipid raft microdomains have been implicated in the assembly and budding of many viruses.[54] Filovirus GP associates with lipid rafts in GP-expressing cells and in virus-infected cells.[17] VP40 also localizes to lipid rafts, predominantly in its oligomeric form.[21] Deletion of the VP40 carboxy-terminus or replacement of prolines 283 and 286 abolishes raft association but does not affect oligomerization.[21] Both of these VP40 mutations result in reduced VLP release, which correlates with the observed decrease in raft association.[21] Thus, raft association appears necessary for efficient particle production. The released virions contain the raft-associated glycolipid GM1,[17] but not the raft-excluded protein transferrin receptor,[17] a finding that implies filoviruses may use lipid rafts as the final gateway for the exit of mature virions, similar to the mechanism used by human immunodeficiency virus type 1,[22] measles virus,[23] Sendai virus[24] and respiratory syncytial virus.[25] Although the precise way in which filoviruses use lipid rafts has yet to be determined, the lipid rafts may serve as building platforms for filovirus assembly.

Use of Cellular Transport Pathways

To perform their functions in particle assembly and budding, viral matrix proteins must make use of the cellular transport pathways.

In fact, the matrix proteins of filoviruses and rhabdoviruses, as well as the Gag proteins of several retroviruses, contain small, conserved motifs that are recognized by cellular proteins involved in protein-sorting and transport.[33,58,59] These motifs are called "late domains" to reflect their function late in the viral lifecycle. To date, three late-domain motifs have been described: P(T/S)AP, PPXY, and YXXL.

The PPXY motif interacts with the WW domain of Nedd4-like E3 ubiquitin ligases,[16,26–28] resulting in mono-ubiquitination of the cargo protein and its recognition by the endosomal sorting machinery. The P(T/S)AP motif is recognized by Tsg101,[28–30] a component of the ESCRT-I (endosomal sorting complex required for transport) complex. The YXXL motif binds to two proteins, AP-2[31] and AIP1.[32,33] AP-2 is involved in endocytosis, whereas AIP1 interacts with a component of the ESCRT-III complex.

The Marburg virus VP40 protein contains a PPXY motif near its amino-terminus, whereas the Ebola virus VP40 protein is characterized by overlapping PTAP and PPXY motifs (PTAPPEY at amino acids 7–13).[16,34] Extensive *in vitro* characterization of VP40 proteins containing mutations or deletions in their late domains has indicated that both domains play a role in virus budding.[6,16,35,36] However, reverse genetics allowed the artificial generation of Ebola viruses with mutations in one or both late-domain motifs.[55] The resultant mutant viruses were only mildly to moderately attenuated in cell culture, demonstrating that the late domains are not absolutely required for Ebola virus replication in cell culture. However, they are likely needed to achieve optimal replication efficiencies.

Many viruses contain two late-domain motifs separated by only a few amino acids (the Ebola virus VP40 protein is an exception in that the late domains overlap). While the late domains were originally thought to be functionally interchangeable and thus redundant, more recent data suggest that both motifs are required for efficient particle formation.[58] The two late domains may trigger sequential steps in which recruitment of Nedd4-like ubiquitin ligases (mediated by the PPXX motif) results in the ubiquitination of the cargo protein. The ubiquitinated cargo protein may then have an increased affinity for

Tsg101 (mediated by the P(T/S)AP) motif), as has been shown for Gag.[37] This hypothesis is supported by the finding that Ebola VP40 is ubiquitinated by Nedd4-like ubiquitin ligases;[38] however, it remains to be seen whether this modification affects its interaction with Tsg101.

Nucleocapsid Assembly

In contrast to the extensive work conducted on the budding process, research on nucleocapsid formation has been limited to date. Filovirus nucleocapsids are thought to consist of viral RNA and the NP, VP30, VP35, and L (and possibly also the VP24) proteins. Three of these proteins (NP, VP35 and L) are necessary and sufficient for replication of a monocistronic genome-like Marburg virus replicon,[39,40] while VP30 acts as a transcription regulator.[40,41] By contrast, all four of these proteins are required for the replication of Ebola virus genome-like replicons.[40] All four proteins are also essential for the generation of Ebola virus[42,43] and Marburg[56] virus from cloned cDNAs ("reverse genetics"). An earlier study demonstrated that VP24 was not completely removed from purified nucleocapsids under isotonic conditions,[44] suggesting that it could also be a constituent of the nucleocapsids, although it is not required for transcription and replication of the viral genome.

Expression of Marburg virus NP in insect cells leads to the formation of NP-RNA complexes,[57] as had been reported with the rabies virus and measles virus NP proteins.[45–47] Purified NP-RNA complexes of Marburg virus show loose coils that differ from authentic Marburg virus nucleocapsids. By contrast, NP-RNA complexes of rabies and measles viruses resemble those found in infected cells. Therefore, additional components are necessary for the formation of Marburg virus nucleocapsids.

Electron microscopic[48] and tomographic studies[49] identified the components of filovirus nucleocapsids (Fig. 5). Simultaneous expression of Ebola NP, VP35, and VP24 in mammalian cells results in the formation of nucleocapsid-like structures indistinguishable from authentic nucleocapsids.[48] Of the three proteins, VP24 seems to be a minor component.[48] NP and VP35 appear to build up the framework of the

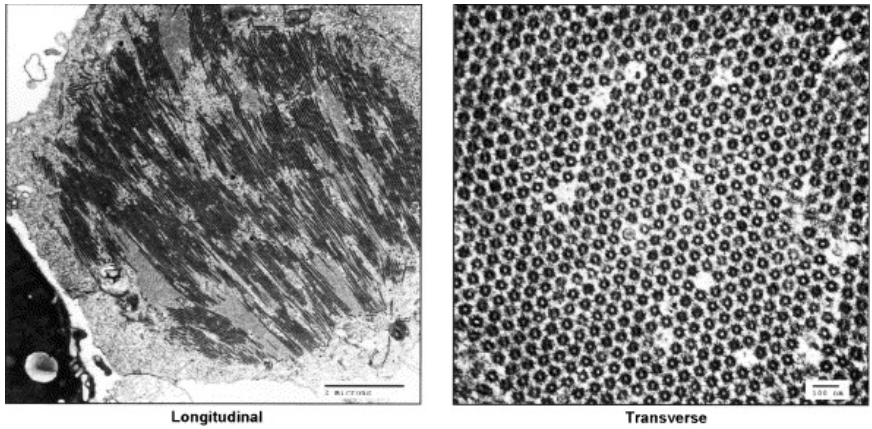


Fig. 5. Formation of nucleocapsid-like structures. Coexpression of NP, VP35, and VP24 is necessary and sufficient for the formation of nucleocapsid-like structures. (Courtesy of Huang *et al.*, 2002.)

nucleocapsids, while VP24 may play a transient catalytic role during nucleocapsid formation or serve as molecular glue between NP and VP35.

Electron microscopic studies have also revealed nucleocapsid aggregation in the cytoplasm at an appreciable distance from the cell surface.[48] Once the newly synthesized genomic RNA is encapsidated in the cytoplasm, the mature nucleocapsids must be transported to the plasma membrane for virus budding. As discussed above, the VP40 protein may be involved in this step. This assumption is supported by the findings that NP and VP40 colocalize in cells expressing both proteins, and that NP expression increases the efficiency of VP40-VLP particle release.[50,51] Moreover, VP40 interacts with VP35.[10] In filovirus-infected cells, however, direct interactions between nucleocapsids and VP40 have not been demonstrated. GP may also be involved in this transport process since its cytoplasmic tail interacts with VP40.

Conclusion

Although the detailed mechanisms of filovirus genesis are still unknown, much has been learned about the individual steps of the assembly and

budding processes. Since diverse viruses seem to exploit common cellular pathways and machinery, knowledge of filovirus assembly and budding may aid in the identification and characterization of other commonly used pathways and stimulate research in other areas. Even more importantly, knowledge of specific interactions among filoviral proteins may prove invaluable in identifying targets for antivirals to Ebola and Marburg viruses.

Acknowledgments

We thank the members of our laboratory who contributed to the original work presented in this review, Yuko Kawaoka for an illustration, Gabriel Neumann for critical reading of the manuscript, and Susan Watson for editing the manuscript. This work was supported by Core Research for Evolutional Science and Technology (CREST) grants from the Japan Science and Technology Corporation (JST), Japan.

References

1. Sanchez A, Khan AS, Zaki SR, *et al.* (2001) Filoviridae: Marburg and Ebola viruses. In DM Knipe & PM Howley (eds), *Fields Virology*, 4th ed, pp. 1279–1304. Lippincott/The Williams & Wilkins Co., Philadelphia, Pa.
2. Geisbert TW, Jahrling PB. (1995) Differentiation of filoviruses by electron microscopy. *Virus Res* **39**: 129–150.
3. Schmitt AP, Lamb RA. (2004) Escaping from the cell: assembly and budding of negative-strand RNA viruses. *Curr Top Microbiol Immunol* **283**: 145–196.
4. Feldmann H, Klenk HD. (1996) Marburg and Ebola viruses. In Maramorosh K, Murphy FA, Shatkin AJ (eds), *Advances in virus research*, vol. 47, pp. 1–52. Academia Press, Inc.
5. Garoff H, Hewson R, Opstelten DJ. (1998) Virus maturation by budding. *Microbiol. Mol Biol Rev* **62**: 1171–1190.
6. Jasenosky LD, Neumann G, Lukashevich I, Kawaoka Y. (2001) Ebola virus VP40-induced particle formation and association with the lipid bilayer. *J Virol* **75**: 5205–5214.
7. Ruigrok RW, Schoehn G, Dessen A, *et al.* (2000) Structural characterization and membrane binding properties of the matrix protein VP40 of Ebola virus. *J Mol Biol* **300**: 103–112.

8. Kolesnikova L, Bugany H, Klenk HD, Becker S. (2002) VP40, the matrix protein of Marburg virus, is associated with membranes of the late endosomal compartment. *J Virol* **76**: 1825–1838.
9. Noda T, Sagara H, Suzuki E, *et al.* (2002) Ebola virus VP40 drives the formation of virus-like filamentous particles along with GP. *J Virol* **76**: 4855–4865.
10. Johnson RF, McCarthy SE, *et al.* (2006) Ebola virus VP35-VP40 interaction is sufficient for packaging 3E-5E minigenome RNA into virus-like particles. *J Virol* **80**: 5135–5144.
11. Dessen A, Volchkov V, Dolnik O, *et al.* (2000) Crystal structure of the matrix protein VP40 from Ebola virus. *EMBO J* **19**: 4228–4236.
12. Scianimanico S, Schoehn G, Timmins J, *et al.* (2000) Membrane association induces a conformational change in the Ebola virus matrix protein. *EMBO J* **19**: 6732–6741.
13. Timmins J, Schoehn G, Kohlhaas C, *et al.* (2003) Oligomerization and polymerization of the filovirus matrix protein VP40. *Virology* **312**: 359–368.
14. Hoenen T, Volchkov V, Kolesnikova L, *et al.* (2005) VP40 octamers are essential for Ebola virus replication. *J Virol* **79**: 1898–1905.
15. Gomis-Ruth FX, Dessen A, Timmins J, *et al.* (2003) The matrix protein VP40 from Ebola virus octamerizes into pore-like structures with specific RNA binding properties. *Structure (Camb)* **11**: 423–433.
16. Harty RN, Brown ME, Wang G, *et al.* (2000) A PPxY motif within the VP40 protein of Ebola virus interacts physically and functionally with a ubiquitin ligase: implications for filovirus budding. *Proc Natl Acad Sci, USA* **97**: 13871–13876.
17. Bavari S, Bosio CM, Wiegand E, *et al.* (2002) Lipid raft microdomains: a gateway for compartmentalized trafficking of Ebola and Marburg viruses. *J Exp Med* **195**: 593–602.
18. Swenson DL, Warfield KL, Kuehl K, *et al.* (2004) Generation of Marburg virus-like particles by co-expression of glycoprotein and matrix protein. *FEMS Immunol Med Microbiol* **40**: 27–31.
19. Volchkov VE, Volchkova VA, Slenczka W, *et al.* (1998) Release of viral glycoproteins during Ebola virus infection. *Virology* **245**: 110–19.
20. Watanabe S, Watanabe T, Noda T, *et al.* (2004) Production of novel ebola virus-like particles from cDNAs: an alternative to Ebola virus generation by reverse genetics. *J Virol* **78**: 999–1005.
21. Panchal RG, Ruthel G, Kenny TA, *et al.* (2003) *In vivo* oligomerization and raft localization of Ebola virus protein VP40 during vesicular budding. *Proc Natl Acad Sci USA* **100**: 15936–15941.

22. Campbell SM, Crowe SM, Mak J. (2001) Lipid rafts and HIV-1: from viral entry to assembly of progeny virions. *J Clin Virol* **22**: 217–227.
23. Manie SN, Debreyne S, Vincent S, Gerlier D. (2000) Measles virus structural components are enriched into lipid raft microdomains: a potential cellular location for virus assembly. *J Virol* **74**: 305–311.
24. Ali A, Nayak DP. (2000) Assembly of Sendai virus: M protein interacts with F and HN proteins and with the cytoplasmic tail and transmembrane domain of F protein. *Virology* **276**: 289–303.
25. Brown G, Rixon HW, Sugrue RJ. (2002) Respiratory syncytial virus assembly occurs in GM1-rich regions of the host-cell membrane and alters the cellular distribution of tyrosine phosphorylated caveolin-1. *J Gen Virol* **83**: 1841–1850.
26. Harty RN, Brown ME, McGettigan JP, *et al.* (2001) Rhabdoviruses and the cellular ubiquitin-proteasome system: a budding interaction. *J Virol* **75**: 10623–10629.
27. Strack B, Calistri A, Accola MA, *et al.* (2000) A role for ubiquitin ligase recruitment in retrovirus release. *Proc Natl Acad Sci USA* **97**: 13063–13068.
28. VerPlank L, Bouamr F, LaGrassa TJ, *et al.* (2001) Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). *Proc Natl Acad Sci USA* **98**: 7724–7729.
29. Garrus JE, von Schwedler UK, Pornillos OW, *et al.* (2001) Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **107**: 55–65.
30. Martin-Serrano J, Zang T, Bieniasz PD. (2001) HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nat Med* **7**: 1313–1319.
31. Puffer BA, Watkins SC, Montelaro RC. (1998) Equine infectious anemia virus Gag polyprotein late domain specifically recruits cellular AP-2 adapter protein complexes during virion assembly. *J Virol* **72**: 10218–10221.
32. Strack B, Calistri A, Craig S, *et al.* (2003) AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell* **114**: 689–699.
33. von Schwedler UK, Stuchell M, Muller B, *et al.* (2003) The protein network of HIV budding. *Cell* **114**: 701–713.
34. Jasenosky LD, Kawaoka Y. (2004) Filovirus budding. *Virus Res* **106**: 181–188.
35. Licata JM, Simpson-Holley M, Wright NT, *et al.* (2003) Overlapping motifs (PTAP and PPEY) within the Ebola virus VP40 protein function

independently as late budding domains: involvement of host proteins TSG101 and VPS-4. *J Virol* **77**: 1812–1819.
36. Timmins J, Scianimanico S, Schoehn G, Weissenhorn W. (2001) Vesicular release of ebola virus matrix protein VP40. *Virology* **283**: 1–6.
37. Pornillos O, Alam SL, Davis DR, Sundquist WI. (2002) Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein. *Nat Struct Biol* **9**: 812–817.
38. Martin-Serrano J, Perez-Caballero D, Bieniasz PD. (2004) Context-dependent effects of L domains and ubiquitination on viral budding. *J Virol* **78**: 5554–5563.
39. Muhlberger E, Lotfering B, Klenk HD, Becker S. (1998) Three of the four nucleocapsid proteins of Marburg virus, NP, VP35, and L, are sufficient to mediate replication and transcription of Marburg virus-specific monocistronic minigenomes. *J Virol* **72**: 8756–8764.
40. Muhlberger E, Weik M, Volchkov VE, *et al.* (1999) Comparison of the transcription and replication strategies of Marburg virus and Ebola virus by using artificial replication systems. *J Virol* **73**: 2333–2342.
41. Weik M, Modrof J, Klenk HD, *et al.* (2002) Ebola virus VP30-mediated transcription is regulated by RNA secondary structure formation. *J Virol* **76**: 8532–8539.
42. Volchkov VE, Volchkova VA, Muhlberger E, *et al.* (2001) Recovery of infectious Ebola virus from complementary DNA: RNA editing of the GP gene and viral cytotoxicity. *Science* **291**: 1965–1969.
43. Neumann G, Feldmann H, Watanabe S, *et al.* (2002) Reverse genetics demonstrates that proteolytic processing of the Ebola virus glycoprotein is not essential for replication in cell culture. *J Virol* **76**: 406–410.
44. Kiley MP, Cox NJ, Elliott LH, *et al.* (1988) Physicochemical properties of Marburg virus: evidence for three distinct virus strains and their relationship to Ebola virus. *J Gen Virol* **69**: 1957–1967.
45. Iseni F, Barge A, Baudin F, *et al.* (1998) Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. *J Gen Virol* **79**: 2909–2919.
46. Prehaud C, Harris RD, Fulop V, *et al.* (1990) Expression, characterization, and purification of a phosphorylated rabies nucleoprotein synthesized in insect cells by baculovirus vectors. *Virology* **178**: 486–497.
47. Fooks AR, Stephenson JR, Warnes A, *et al.* (1993) Measles virus nucleocapsid protein expressed in insect cells assembles into nucleocapsid-like structures. *J Gen Virol* **74**: 1439–1444.

48. Huang Y, Xu L, Sun Y, Nabel GJ. (2002) The Assembly of Ebola Virus Nucleocapsid Requires Virion-Associated Proteins 35 and 24 and Posttranslational Modification of Nucleoprotein. *Mol Cell* **10**: 307–316.
49. Noda T, Aoyama K, Sagara H, *et al.* (2005) Nucleocapsid-like structures of Ebola virus reconstructed using electron tomography. *J Vet Med Sci* **67**: 325–328.
50. Licata JM, Johnson RF, Han Z, Harty RN. (2004) Contribution of Ebola virus glycoprotein, nucleoprotein, and VP24 to budding of VP40 virus-like particles. *J Virol* **78**: 7344–7351.
51. Kallstrom G, Warfield KL, Swenson DL, *et al.* (2005) Analysis of Ebola virus and VLP release using immunocapture assay. *J Virol Methods* **127**: 1–9.
52. Noda T, Ebihara H, Muramoto Y, *et al.* (2006) Assembly and budding of Ebolavirus. *PLoS Pathology* 2: e99.
53. Schnittler HJ, Mahner F, Drenckhahn D, *et al.* (1993) Replication of Marburg virus in human endothelial cells. A possible mechanism for the development of viral hemorrhagic disease. *J Clin Invest* **91**: 1301–1309.
54. Nayak DP, Barman S. (2002) Role of lipid rafts in virus assembly and budding. *Adv Virus Res* **58**: 1–28.
55. Neumann E, Ebihara H, Takada A, *et al.* (2005) Ebola virus VP40 late domains are not essential for viral replication in cell culture. *J Virol* **79**: 10300–10307
56. Enterlein S, Volchkov V, Weik M, *et al.* (2006) Rescue of recombinant Marburg virus from cDNA is dependent on nucleocapsid protein VP30. **80**: 1038–1043.
57. Mavrakis M, Kolesnikova L, Schoehn G, *et al.* (2002) Morphology of Marburg virus NP-RNA. *Virology* **296**: 300–307.
58. Demirov DG, Freed EO. (2004) Retrovirus budding. *Virus Res* **106**: 87–102.
59. Freed EO. (2002) Viral late domains. *J Virol* **76**: 4679–4687.

This page intentionally left blank

Chapter 13

Challenges in Designing HIV Env Immunogens for Developing a Vaccine

Indresh K. Srivastava and R. Holland Cheng†*

Summary

HIV continues to be a major health problem worldwide; however, the situation is particularly serious in Asian and Sub-Saharan countries. Development of an effective HIV vaccine could help to reduce the severity of the disease and prevent infection. Over the last two decades significant efforts have been made toward inducing potent humoral and cellular immune responses by vaccination; however, it appears that either antibodies or CTL may not be sufficient alone for the induction of sterilizing immunity or long-term control of viral replication. Therefore, it is generally believed that both humoral and cellular responses will be needed for an effective HIV vaccine. It has been shown in passive transfer experiments using broadly neutralizing monoclonal antibodies (mAb) such as b12, 2F5, and 2G12 that these mAbs either alone or in combination are effective in conferring

* Associate Director, Vaccines Research, Novartis Vaccines and Diagnostics, Inc., Mail Stop 4.3, 4560 Horton Street, Emeryville, CA 94608 (Address all correspondence to Indresh K. Srivastava). Phone: 510-923-5485, Fax: 510-923-2586, Email: indresh.srivastava@novartis.com.

† Professor of Molecular and Cellular Biology, Advanced Microscopy and Proteomics, University of California, Briggs Hall, Davis, CA 95616, USA.
Karolinska Institute, Structural Virology, Novum, Hälsovägen 7, Huddinge, SE 14157, Sweden.

protection against challenge infection to rhesus macaques. However, for the development of an effective vaccine, it has been a challenge to induce protective antibodies of similar specificities by vaccination. Therefore, efforts are being made by different groups to design an Env immunogen that may be more effective compared to the existing immunogens in inducing potent neutralizing and protective antibody responses by: i) optimizing existing Env structure to enhance the exposure of functional epitopes to focus the responses to these epitopes; ii) obtaining structural information on HIV Env, and using this information for structure-based novel immunogen design; and iii) identifying novel functional epitopes, and designing strategies to incorporate them in potential vaccines. Once potent HIV Env structures have been identified, their effectiveness may be enhanced through the use of adjuvants, delivery systems, and prime and boost strategies to improve the quality and magnitude of neutralizing responses.

Introduction

AIDS continues to be a major health problem throughout the world, with approximately 40 million cases and 20 million deaths recorded so far. In certain parts of the world, such as Sub-Saharan Africa, the prevalence of human immunodeficiency virus (HIV) in the population is estimated to be as high as 35%.[1] If current infection rates and the absence of affordable treatments continue, 60% of the current adolescent population in that region will not live to the age of 60.[1] In the United States today, an estimated 950,000 individuals are living with HIV, and 40,000 to 80,000 new infections occur each year.[1] Moreover, the situation is continuously deteriorating as a result of the rapid emergence of drug resistance against most of the effective antivirals. Therefore, there is an urgent need for an effective anti-HIV vaccine that may be used either alone as a prophylactic vaccine or in conjunction with anti-viral drugs as a therapeutic treatment.

Based on preclinical and early clinical studies, it was concluded that neutralizing antibody responses induced by monomeric HIV Env are not potent enough to protect against the HIV infection.[2,3] Therefore the dogma shifted from the induction of neutralizing

antibodies to the induction of cellular responses. Previous studies used the gag to show the correlation between strong cytotoxic T-cell activity and reduced viral load. However, more recent data obtained in a rhesus macaque challenge model suggest that cellular responses focused on a single epitope may not be sufficient either to effectively control viral replication.[4–7] It remains to be seen if increasing the breadth of CTL responses will have a positive impact on the outcome of challenge infection. Hence, it is now believed that humoral, cellular, and T-helper cell responses, both at peripheral and mucosal sites, are needed for broad and durable protection against HIV. Conceptually, antibodies would serve as a first line of defense by completely preventing infection (Fig. 1A) or reducing the virus inoculum, whereas cellular responses would facilitate the clearance of HIV-infected $CD4^+$ T-cells that escape antibody-mediated neutralization

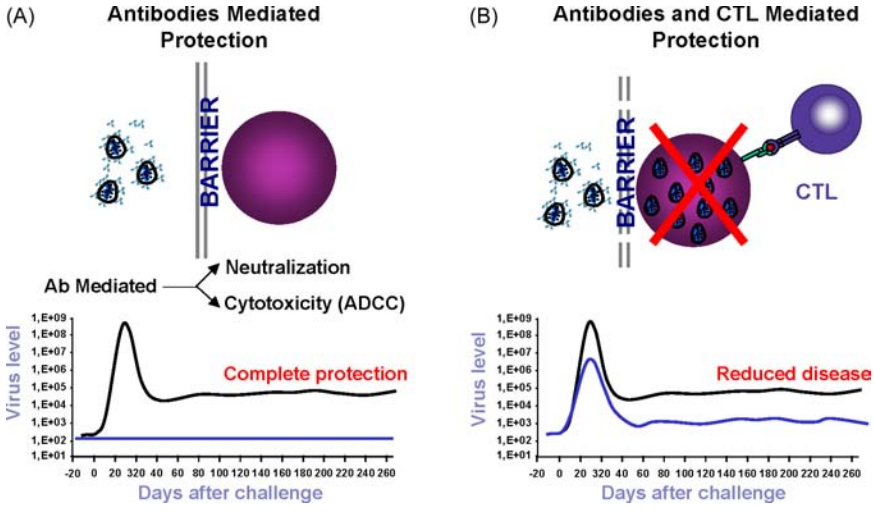


Fig. 1. Two potential outcomes for a preventative HIV vaccine. In the graphs, the black line shows time course of viremia in the absence of vaccination. In panel A, the blue line indicates the virologic outcome in the case of antibody-mediated (neutralization and ADCC) complete protection. In panel B, the blue line indicates the outcome in the case of a vaccine that induces partial protection and reduces the disease where both humoral and cellular responses participated.

(Fig. 1B) and would thereby reduce the severity of the disease. However, Mascola and colleagues have shown that the induction of cellular responses using DNA vaccines in conjunction with passive transfer of neutralizing antibodies did not offer additional protection compared to the protection afforded by passive transfer of neutralizing antibodies alone.[8] Further studies will be needed to demonstrate the advantage of inducing potent cellular and humoral responses by an effective HIV vaccine for inducing the sterilizing immunity or effectively controling the level of viral replication.

For the induction of neutralizing antibody responses, HIV Env glycoprotein is the major antigenic target. However it has proven difficult to induce neutralizing antibody responses of appropriate specificity against diverse primary HIV-1 strains utilizing monomeric HIV Env (i.e. gp120) glycoprotein. Therefore, a major challenge is to develop novel strategies to identify and expose critical epitopes that may be the target for inducing such broadly cross-reactive neutralizing antibodies with protective efficacy. This focus of this review is: a) to capture the progress made in structural analysis of critical neutralizing antibodies, and b) to present strategies to design more potent immunogens that may induce neutralizing antibodies of protective efficacy.

Role of Antibodies in Protection Against HIV

Antibodies serve as a correlate of protection against some of the bacterial[9,10] and viral pathogens,[11] therefore targeting humoral responses has been critical for developing prophylactic vaccines against these targets. In the case of HIV, it has been a daunting task to make an effective vaccine due to a lack of correlates of protection. During the natural course of HIV infection or immunization with purified protein, a very strong humoral response is induced. However, most of these antibodies are binding antibodies, and only a fraction of these antibodies have neutralizing properties. These antibodies are termed as "neutralizing antibodies" (nAb) because they neutralize the infectious virus, and therefore prevent its entry to the cells or reduce the level of inoculum, thereby reducing the severity of the disease.[12]

However, only a proportion of all neutralizing antibodies have protective efficacy. Until recently, it was generally believed that not all Env spikes need to be engaged by these nAbs for them to be efficient in controlling or preventing viral replication. However, Yang and colleagues demonstrated that virus neutralization requires essentially all of the functional trimers to be occupied by at least one antibody. Their model applies to antibodies differing in neutralizing potency and to virus isolates with different neutralization sensitivities.[13] However, recent data suggest that in addition to nAbs, there is another class of antibodies that exert their protective efficacy by inducing antibody dependent cellular cytotoxicity (ADCC); i.e. they direct the killing of infected cells through recognition of viral proteins on cell surfaces.

Possible Steps for Immune Intervention

There are several possible steps during the virus lifecycle (pre- and post-attachment events) where immune intervention is possible and may prevent the infection or control viral replication, as illustrated in Fig 2. Since binding of HIV to its receptor (CD4) and co-receptor (e.g. CCR5) is critical for viral entry in new $CD4^+$ T-cells, a logical and attractive strategy could be to prevent the interaction of the HIV virion with its cellular receptor or co-receptor. Other possible and attractive strategies could be to target other important post-attachment events such as: a) the formation of a coiled-coil structure leading to virus/host membrane fusion[14–17]; b) primary un-coating of the virus in the cytoplasm; c) transcription or virus assembly; and d) budding of the new virion. For HIV-1, inhibition of receptor binding has been seen for different types of antibodies, including: i) those that target carbohydrate of the virus glycoprotein[18,19]; ii) those that target different ***adhesion molecules***; iii) those specific for various regions of gp120 such as the ***CD4 binding domain***, the ***CD4-induced (CD4i) epitope*** (located in the bridging sheet of gp120 that is created or exposed when gp120 interacts with CD4),[20] and the ***variable loops***, V2 and V3; iv) those that target the ***co-receptor binding site***[21]; and (v) another class of mAbs such as 2F5, 4E10, or Z13 that may

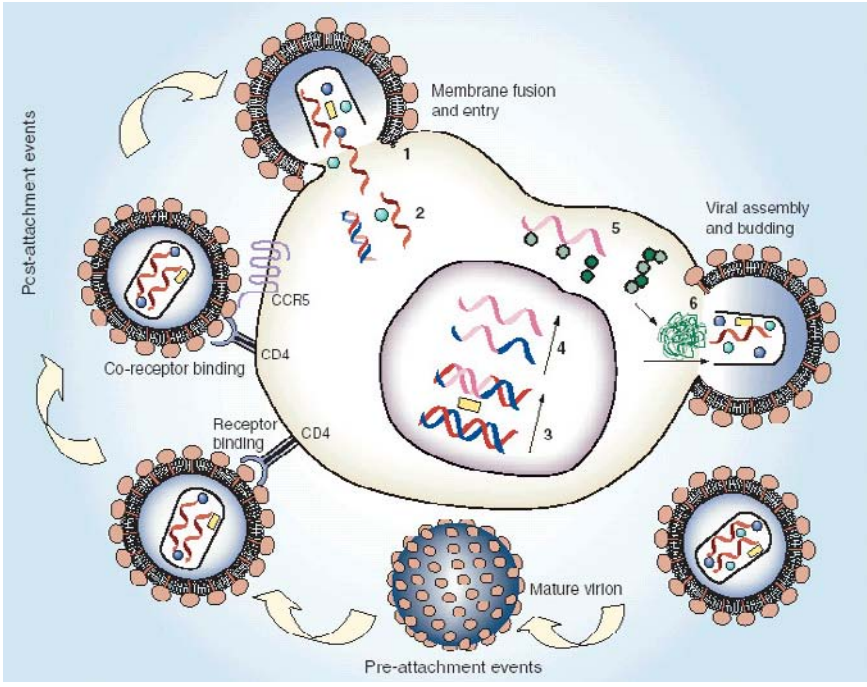


Fig. 2. Lifecycle of HIV-1 virus and potential sites of immune intervention. Step 1: Pre-attachment events: block the binding of virion to its receptor (i.e. CD4); complement mediated lysis and aggregation of infectious virion. Step 2: Post attachment events — involves prevention of binding to co-receptor and prevention of virus/cell fusion. In addition, these neutralizing antibodies may also interfere in downstream events such as: (2-4) uncoating, integration, and reverse transcription after entry; (5) processing and assembly; and (6) maturation and budding of virus particles.

interfere with gp41 coiled-coil formation or fusion of the virus/ host cell membrane. Some of these epitopes in context to Env trimer are presented in Fig. 3. Another mechanism by which antibodies can mediate protection is aggregation of virus particles, thereby reducing the number of infectious virus particles and rendering the virus more susceptible to phagocytosis and subsequent destruction.[22]

Another mechanism by which binding antibodies may exert a protective function is complement-mediated ADCC, where antibodies

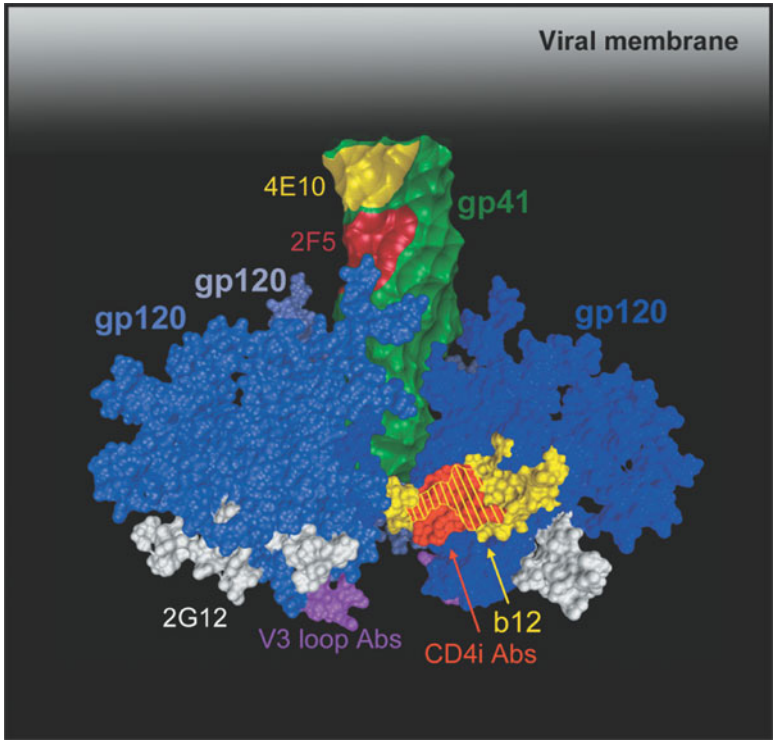


Fig. 3. Neutralizing epitopes in context to trimers. (Figure is adapated from Burton *et al.* (2004) *Nature Immunol* **5**: 233–236.)

act as a bridge between Fc receptor (FcR)-bearing cells and HIV-1-infected cells or other cells that have passively absorbed HIV-1 Env onto their surface. The killing of virus-infected cells by ADCC involves cytotoxic cells (such as NK cells) and may contribute toward the elimination of the virus.[23–26] Alsmadi and Tilley[27] studied the ability of human and chimpanzee mAbs directed against cluster II overlapping epitopes of gp41, CD4 binding site, V3 loop, and C5 domain of gp120 for their ability to induce ADCC. They demonstrated that mAbs directed against conserved epitopes generally exhibited ADCC activity against a broader range of HIV-1 strains than those directed against variable epitopes. Furthermore, it appears that many if not most neutralizing antibodies of high affinity and of the immunoglobulin G1 subclass (IgG1) are capable of exerting ADCC. In a

multi-centered AIDS cohort study, early ADCC responses in patients were associated with higher numbers of $CD4^+$ T-cells and the absence of lymphadenopathy during the first two years of follow up.[28] In two other studies, ADCC responses correlated inversely with plasma viral load and also with $CD4^+$ T-cell counts.[24,29] The presence of neutralizing and ADCC activity in children born to HIV-infected mothers correlated better with the clinical outcome.[30] Finally, it has been shown that ADCC antibodies were present in cervicovaginal fluids in HIV-1 infected women,[31] supporting the hypothesis that this form of immunity can contribute in protection against HIV-1 at the site of virus entry. In a recent study, Gomez-Roman *et al.* have shown that priming with replicating adenovirus type 5 host range mutant-SIV recombinants, followed by boosting with SIV gp120, elicited potent ADCC activity that correlated with protection against the mucosal challenge infection with pathogenic SIVmac251 in rhesus macaques.[32] This is the first study that demonstrates a good correlation between the *in vitro* ADCC activity and *in vivo* reduced viremia; however, further studies will be needed to establish the relationship between the induction of ADCC and protection against HIV. The identification of "ADCC epitopes" may help in designing strategies on means to present and enhance the potency of these epitopes to prevent or control the viral infection.

Neutralizing Epitopes Relevant for Vaccine Development

As mentioned above, Env is the most important target for induction of antibody responses against HIV. In addition, it has several known CTL and helper T-cell epitopes that may provide targets for an effective anti-HIV vaccine. The *env* gene is expressed during the late phase of viral transcription as the gp160 precursor protein. Translation of the precursor protein is dependent on the viral Rev protein, which binds to the rev responsive element (RRE) in Env mRNA and mediates its nuclear export.[33] During the maturation of the virus, the gp160 is proteolytically processed by cellular serine proteases to yield: i) membrane-spanning domain termed gp41 and ii) an extracellular

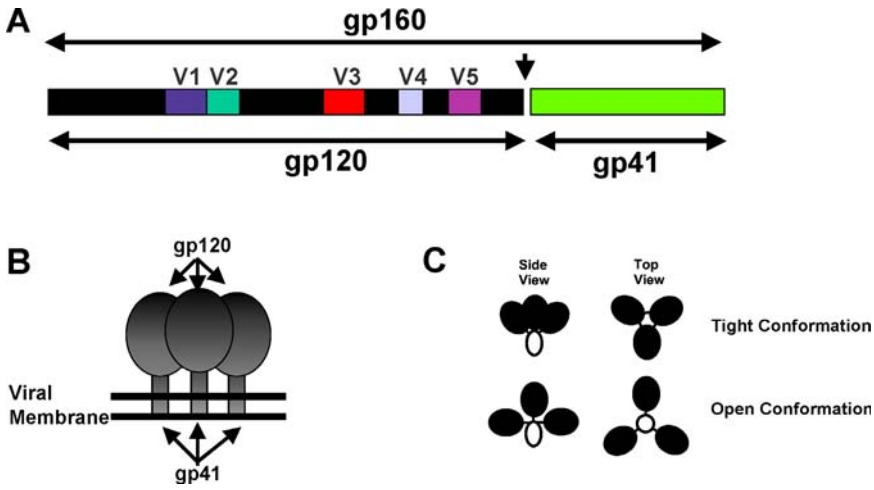


Fig. 4. Linear map of HIV Env (SF162) indicating various conserved and variable domains (Panel A), and large numbers of cysteine residues and extensive glycosylation, likely conformation of Env involving gp120 ectodomain, and gp41 on the surface of the virus (Panel B). Also shown is the schematic of trimers in open and closed confirmation (Panel C). Further structural studies will be needed to demonstrate if the native trimers are in open or closed conformation.

domain termed gp120 (Fig. 4A). Env proteins become heavily glycosylated during their passage through the Golgi apparatus. gp120 and gp41 are non-covalently associated on the viral surface to form trimeric spikes (Fig. 4B) that can bind to the CD4 on T-cells, which is the primary receptor of HIV-1. Comparison of the sequences of the *env* gene from different isolates and clades reveals that it has five hypervariable regions and five conserved regions, and a large number of cysteine residues (Fig. 4A). There are differences between the size of the variable loops and the number of glycosylation sites in V1 and V2 loops among virus isolates of different clades, and between early and late primary viruses. This sequence variation in the hypervariable loops is due to nucleotide changes and subsequent accumulation of point mutations resulting in amino acid substitutions. These mutation-induced changes in Env tertiary structure are selected by the

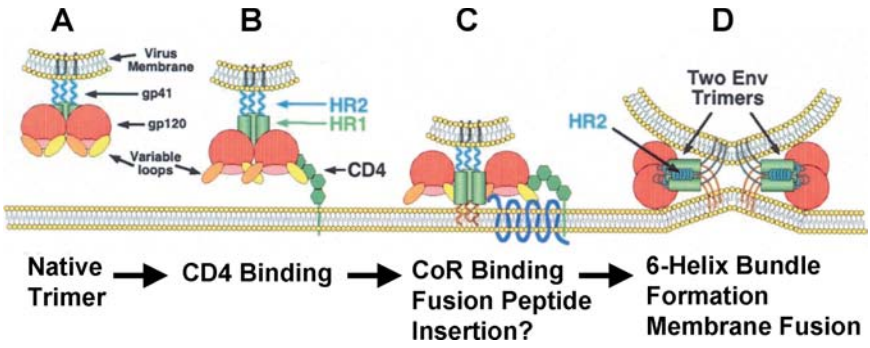


Fig. 5. Receptor and co-receptor mediated entry of virus into CD4+ T-cell. The figure is adapted from Moore and Doms.[34] After CD4 binding, gp120 undergoes a conformational change and exposes the co-receptor binding site on Env (Panel B). The triggered Env binds to a seven-trans-membrane domain co-receptor (third section, CoR). The hydrophobic fusion peptide at the N terminus of gp41 becomes exposed and inserts into the membrane of the cell (Panel C). Whether this results from CD4 binding or co-receptor binding is not known. Co-receptor binding ultimately results in formation of a six-helix bundle in which the helical HR2 domains in each gp41 subunit fold back and pack into grooves on the outside of the triple-stranded HR1 domains (Panel D), bringing the fusion peptide and trans-membrane domain of gp41 (and their associated membranes) into close proximity. It is likely that several Env trimers need to undergo this conformational change in order to form a fusion pore, although here only two trimers are depicted. It is not known whether gp120 remains associated during the fusion process or dissociates from gp41. Although only a single CD4-binding event is shown, multiple CD4-binding events may be needed to activate a single Env trimer.

immune responses of the host and provide a means of immune escape. The receptor and co-receptor dependent entry process of HIV into target cells is depicted in Fig. 5.[34] Interaction of HIV Env with the CD4 is an obligatory step for virus entry into the cell; therefore, as expected, the CD4 binding domain of gp120 is a highly conserved, complex, and conformational dependent region. Hence, the **CD4 binding site of Env (CD4BS)** may be an excellent target for immune intervention. Consistent with this hypothesis, many mAbs

against the CD4 BS have been developed that can neutralize T-cell line-adapted isolates,[35,36] suggesting that these epitopes are relatively well exposed on the virion surface.[37] The most potent and well-characterized monoclonal antibody against the CD4BS, b12,[38] neutralizes a broad range of primary isolates and confirms the critical role of the CD4BS in HIV-1 infection.[39–41] Interestingly, other CD4BS monoclonal antibodies such as 559/64D, 15e, F105, b3, and b6 do not neutralize primary isolates.[42,43] The reasons for this discrepancy are not yet understood, but b12 differs from all the other CD4BS antibodies in its sensitivity to V1-V2 loop deletion.[44] It is not known if b12 contacts the V1-V2 loop or if the sensitivity is due to an indirect effect of conformational rearrangements following V1-V2 deletions. High resolution crystal structure of b12 has been solved, and a key feature of the antibody-combining site is the protruding, finger-like long CDR H3 that can penetrate the recessed CD4-binding site of gp120, which is quite effective in neutralizing the virus. A docking model of b12 and gp120 reveals severe structural constraints that explain the extraordinary challenge in eliciting effective neutralizing antibodies similar to b12. The structure, together with mutagenesis studies, provides a rationale for the extensive cross-reactivity of b12 and is a valuable framework for the design of HIV-1 vaccines capable of eliciting b12-like activity.[45]

After primary attachment of virus to the T-cell surface, gp120 interacts with chemokine receptors CCR5 or CXCR4, which are the most common cellular co-receptors for HIV-1. The interaction of Env to CD4 induces the conformational change in gp120 and gp41 (Fig. 5B). As a result, the co-receptor binding site and fusogenic region of gp41 are displayed (Fig. 5C), leading to the fusion of viral and cellular membranes (Fig. 5D) and, release of the viral core particles in to the cytoplasm of the cell. Therefore ***CD4-inducible epitopes of Env*** may represent another target for immune intervention. Antibodies that show reactivity toward HIV Env when it is complexed with soluble CD4 (sCD4) were found in HIV-1-infected individuals,[46] suggesting that these epitopes are immunogenic. Several such human mAbs have been identified, including 17b, 48D, CG10, 23E, and X5.[47–49] The region recognized by these anti-CD4i monoclonal

antibodies has now been mapped to β-strands in the V1/V2 stem and the C4 region of gp120. These regions are involved in binding of Env to chemokine receptors.[49–52] Surprisingly, 17b and other CD4i mAbs in their complete IgG form do not neutralize most primary isolates.[49] In contrast, Fab and single-chain fragments of these anti-CD4i mAbs neutralize primary isolates,[48,53] suggesting that there may be space constraints between the CD4i epitope on gp120 and the target cell membrane.[54] Therefore, the entire IgG molecule may not be capable of neutralization due to steric hindrance.[55,56] Further studies are needed to confirm the accessibility of these conformational epitopes for targeting them for vaccine application.

Contrary to earlier beliefs that epitopes in the variable domains are isolate-specific (thus may not be appropriate targets for vaccine application), recent structural studies suggest that **V3 loop does contain conserved functional epitopes**, which may be targeted by vaccines. The V3 loop is immunogenic and anti-V3 antibodies are induced early during infection and after immunization.[3,57–59] However, a large proportion of these antibodies are directed against linear epitopes in the V3 loop, which serve as decoys for directing immune responses away from conserved V3 regions. These antibodies neutralize homologous isolates, but they have little or no neutralizing activity against diverse primary isolates.[43,60] However, broadly neutralizing anti-V3 antibodies directed against the conserved conformational epitopes have been described.[61] The most broadly reactive of these neutralizing anti-V3 mAbs (such as 447-52D, 19b, and 2182) can neutralize a large proportion of clade B primary isolates[39,62–64] and also have been shown to neutralize viruses from clades A, F, C, and G,[65] suggesting that the epitopes recognized by these mAbs are conserved across clades.

Structural studies have shown that V3-loop has some constant features, such as a relatively fixed size (30–35 aa), a conserved type-II turn at its crown, a disulfide bond at its base, and a net positive charge.[66,67] These features are required in the V3 loop in order for it to interact with the chemokine receptor.[68,69] It has been shown by Cao *et al.* that deletion of V3 loop renders the virus non-infectious, suggesting that V3 loop is essential.[70] Recently, structural studies have demonstrated that V3 loops in R5 viruses are homologous to the β-hairpin structures

in CC chemokines (CCL3 [MIP-1α], CCL4 [MIP-1 β], and CCL5 [RANTES]), but homologous to the CXC chemokine CXCL12 (SDF-1) in X4 viruses.[71] In addition, Yonezawa *et al.* have replaced the V3 loop of the X4 virus with a 43-amino acid region of SDF-1, which includes the β-hairpin, and demonstrated that infectivity was maintained.[72] Together, these data suggest a critical role for the V3 region in virus infectivity and provide a potential rationale for designing immunogens that induce antibodies to the conserved V3 motif. However, these concepts need to be evaluated in pre-clinical studies for their ability to induce cross-reactive neutralizing antibody responses.

Carbohydrate-dependent Epitopes in gp120

HIV Env is heavily glycosylated, with approximately 50% of its mass due to carbohydrates.[73] It has been demonstrated that glycosylation is critical for the Env-CD4 interaction, since non-glycosylated Env protein (env2-3) does not bind to CD4.[74,75] Thus, carbohydrate moieties on Env appear to provide a functional conformation to Env critical for its interaction with CD4. In addition, based on the crystallization studies of gp120, it appears that the exposed face of gp120 is heavily glycosylated (Fig. 4A). Even in context of trimer, molecular modeling studies suggest that the heavily glycosylated part of the molecule is exposed to the immune system.[76,77] These data suggest that extensive glycosylation may shield critical neutralizing epitopes. Therefore, it is not surprising that antibodies recognizing carbohydrate-dependent epitopes on Env are not readily induced during the course of natural infection. The best-studied ***anti-carbohydrate antibody*** with broadly neutralizing activity is ***mAb 2G12***, which targets a cluster of carbohydrate moieties in gp120.[19,78] This mAb has broad neutralizing activity both against T-cell line-adapted and primary HIV-1 isolates.[19,79] However, its reactivity may be limited, as it does not neutralize subtype C isolates.[80] The unconventional configuration of this mAb[81] and the poor immunogenicity of the epitope recognized by 2G12[79] raise questions about the mode of neutralization (since the epitope recognized by this mAb has been localized on the immunologically silent face of gp120) and how to design an

immunogen with optimal exposure of this epitope. Based on oligomeric modeling studies it seems that the immunologically silent face of Env is likely to face toward the target cell membrane, and thus it is conceptually possible that 2G12 may neutralize the virus by preventing the interaction of the virus with target cells.

As with most pathogens, HIV-1 induces a polyclonal antibody response to a wide array of epitopes on different viral proteins. Studies of polyclonal sera have helped to identify the specificities of antibodies that are associated with protection. It has been shown that sera collected from some HIV-infected individuals neutralize primary isolates (using CCR5 as a co-receptor, R5).[82,83] In addition, several investigators have generated potent neutralizing monoclonal antibodies from the bone marrow of HIV^+ patients against critical functional and conserved epitopes (Table 1), such as the ones in the CD4-binding site (CD4-BS),[38] CD4-inducible epitopes (CD4-i),[46] carbohydrate-dependent epitope,[79] and the epitopes present in variable loops[58,59,84–86] and gp41 regions of Env.[87–89] The relative position of these epitopes compared to the host and target membrane and the complexity of the binding sites is presented in Fig. 3. Furthermore, neutralizing antibodies from the HIV^+ patient's sera can be affinity-purified on a gp120 column,[83,90] suggesting that neutralizing epitopes are present and exposed on gp120. Induction of antibodies to each of these epitopes may ultimately be useful in protecting individuals against HIV-1 infection; however, presentation of each of these epitopes in context to a vaccine is the challenge.

During the past few years, considerable attention has been focused on neutralizing antibodies; therefore, two major points need to be addressed: (i) characterizing the fine specificity of protective antibodies, and (ii) means to elicit these protective antibodies by immunization.

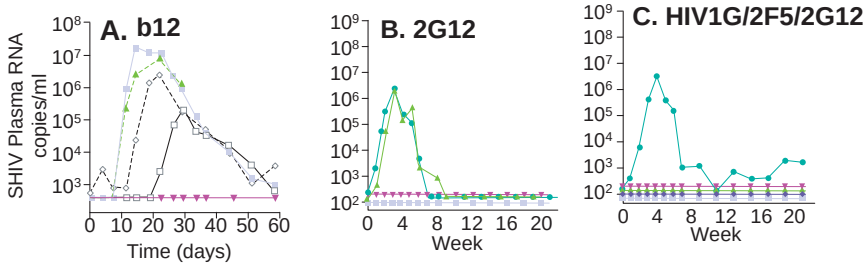
Protective Efficacy of Neutralizing Monoclonal Antibodies in Passive Transfer

It is generally accepted that antibodies are important for protection against HIV infection; however, so far it has been difficult to induce protective antibodies of the appropriate specificity by vaccination. To have a better understanding of the role of neutralizing antibodies

Table 1. Conserved epitopes on the ENV that are targets for developing an HIV vaccine

Env Region	Targeted Epitope	nAbs	Non-nAbs	Characteristics
gp120	CD4-Binding Site (CD4BS)	F105 b12	15e 21h 559/64D 650-D 448D 39.3 b3 b6 830D	These antibodies compete with CD4 for binding to Env. Not all of the CD4BS antibodies neutralize primary isolates.
	CD4-inducible conformational epitope	E51 X5 CG10	17b 48D 23E 49E 21C	Binding of Env to CD4 enhances the exposure of these epitopes. Most of these antibodies neutralize primary isolates as Fab and not as the whole IgG.
	Carbohydrate dependent epitope	2G12		Poorly immunogenic, and binding is dependent upon proper N-linked glycosylation.
gp41	Epitopes in close proximity to viral membrane	2F5 4E10 Z13		These antibodies interfere in membrane fusion and therefore prevent viral entry. To date, these are the most potent neutralizing antibodies identified.
	Cluster I of gp41	Clone 3 246-D		Highly immunogenic epitope, but Clone 3 is the only one of many mAbs specific for this epitope that has neutralizing activity.

in preventing or controlling the HIV infection, the most commonly used approach is to perform passive immunization using neutralizing mAbs in non-human primates. Emini and colleagues have shown that chimpanzees were protected by an anti-gp120 V3 loop-specific



Parren *et al*. (2000); *J Virol* 75: 8340–8347
Mascola *et al*. (2000); *Nature Med* 6: 207–210

Fig. 6. Dose-dependent protection conferred by either single (b12, and 2G12) or a combination of neutralizing antibodies (2F5, and 2G12) to animals in passive transfer experiments as demonstrated Refs. 188 and 189.

antibody[91] against a challenge infection by a TCLA isolate. Subsequently, Conley *et al.* demonstrated the efficacy of mAb 2F5 in partially protecting chimpanzees against intravenous challenge with a primary isolate.[92] The limited availability of non-human primates, particularly chimpanzees, led the investigators to develop an *in vivo* rodent model using severe combined immunodeficient (SCID) mice transplanted with human PBL (hu-PBL-SCID mice) to study the protection afforded by mAbs against HIV-1 infection.[93,94] Using this model it has been shown that passive administration of neutralizing mAbs prior to or shortly after challenge could protect the mice against a challenge infection.[95–97] In this model the antibody concentration needed to protect against the challenge infection *in vivo* was 10 times higher than the concentration needed to neutralize the same isolate *in vitro*.[98] It is interesting, but perhaps not surprising, to observe the differences in the protective efficacy of these neutralizing mAbs *in vivo* and *in vitro*. There are several factors that may influence the protective efficacy of these mAbs, such as dose of the challenge virus and also the growth kinetics of the virus *in vivo* versus *in vitro*. Parren *et al.* demonstrated that higher antibody concentration is required to neutralize primary isolates compared to T-cell adapted isolates.[99] Similar observations were made by Mosier and colleagues

against primary isolates.[100] The use of this model for the evaluation of antibody efficacy is limited due to concerns regarding the route of viral challenge in the mouse model. However, the presence of human cells in the context of the peritoneal and lymph node architecture, complement and phagocytic cells, and the ability to sustain viral infection make it an interesting initial model to test the protective efficacy of passively administered antibodies.

To better understand the qualitative and quantitative characteristics of antibody-mediated protection against HIV-1, Parren and colleagues and Mascola and colleagues used a rhesus macaque challenge model, where animals were passively transfused with neutralizing antibodies and then challenged with the chimeric simian/human immunodeficiency virus (SHIV).[101–103] These viruses contain the Env from HIV-1 and structural proteins from SIV, and are pathogenic in macaques, thereby allowing evaluation of the protective efficacy of antibodies directed against Env. Parren and colleagues passively transferred different concentrations of mAb b12 and then challenged the animals. It was observed that the protection afforded by b12 to rhesus macaques against SHIVSF162P4 is concentration dependent (Fig. 6A). At the concentration of 25 mg/kg of the antibody, all the animals were solidly protected. However, at a 5 mg/ kg level, 50% of the animals were protected. Not surprisingly, at the lowest concentration of 1 mg/kg, none of the animals were protected against the challenge. Mascola and colleagues have demonstrated that a combination of neutralizing antibodies (HIVIG, 2F5, 2G12) was protective against i.v. challenge,[104] where two out of three animals were completely protected while the third animal had a two-log reduction in viral load set point and $CD4^+$ T-cells were also preserved. However, in follow-up experiments it was shown that passive transfer of single antibodies (2F5 or 2G12 or HIVIG) did not protect the animals against the challenge infection because all the animals were infected,[104] suggesting that breadth of immune response may be critical for the protective efficacy. This was contrary to what was observed for b12; however, the challenge virus was different for these studies, and different mAbs were used for passive transfers. Since the primary mode

of transmission for HIV-1 is through mucosal surfaces, efforts have been made to evaluate the efficacy of antibodies for protection against a mucosal challenge.[102] As with systemic challenge, it was demonstrated that complete protection from mucosal challenge was obtained in a majority of animals that received a triple antibody combination (HIVIG/2F5/2G12) (Fig. 6B); however, two out of five animals that received a double combination (2F5/2G12) and two out of four animals that received 2G12 alone were protected against the challenge infection, once again indicating that breadth of the immune response is critical for the protective efficacy (Fig. 6B). To address post-exposure use of neutralizing mAb, Nishimura and colleagues asked the question how soon after virus exposure neutralizing antibodies must be present to block an SIV/HIV chimeric virus infection in pig-tailed macaques.[105] They demonstrated that sterilizing immunity can be achieved in 75% of the animals that received neutralizing IgG six hours after intravenous SIV/HIV chimeric virus inoculation,[105] suggesting that antibodies of appropriate specificity have not only prophylactic but also have therapeutic application.

Ruprecht and colleagues have performed several studies to evaluate the role of neutralizing antibodies in preventing mother-to-infant transmission of virus.[106–108] They demonstrated that passive transfer of a mAb cocktail (F105, 2F5, and 2G12) completely protected pregnant mothers against intravenous SHIV-vpu$^+$ challenge after delivery. The infants subsequently born to these infected mothers who received the mAbs indirectly across the placenta from their mothers and received another dose of antibody cocktail followed by oral challenge with SHIV-vpu$^+$ were also completely protected against the infection.[106] Subsequent experiments have demonstrated that it was possible to protect neonates against a highly pathogenic SHIV 89.6P challenge by passive infusion of a mAb cocktail containing F105, 2F5, 2G12, and b12. Another infusion of the same mAb cocktail an hour after the virus exposure, followed by another dose on day eight, completely protected animals against infection.[106,109] In a recent study, Ferrantelli *et al.*[110] demonstrated complete protection of neonatal rhesus macaques against exposure to pathogenic simian-human immunodeficiency virus (SHIV) by human anti-HIV mAbs. In this study,

eight animals were orally exposed to the pathogenic SHIV and four of them were given post-exposure prophylaxis with three anti-HIV mAbs (2F5, 2G12, and 4E10) at 40 mg/kg dose. All the animals that received mAb therapy were protected from infection ($p = 0.028$) and their plasma, peripheral blood mononuclear cells, and lymph nodes remained free of virus for more than a year. In contrast, all the controls experienced high viral RNA levels and the loss of $CD4^+$ T-cells and died (median survival time 5.5 weeks).

Challenges in Inducing Antibodies of Approriate Specificity with Broadly Neutralizing Activity

During the course of natural infection, HIV triggers antibodies, cytotoxic T-cell (CTL), and $CD4^+$ T-helper immune responses. In general, the primary peak of viremia declines before the appearance of neutralizing antibodies against HIV Env. HIV-infected individuals may generate potent neutralizing antibody responses to autologous isolates, but these responses are slow to develop and take a long time to mature.[3,82,111–116] Interestingly, some long-term non-progressors who remained disease free for more than 10 years after HIV infection developed strong, broadly cross-reactive neutralizing antibody responses, which may have contributed toward their ability to control infection.[117–120] The induction of antibody responses by monomeric Env-based protein subunit vaccines is initially modest and requires multiple boosts to induce strong responses. In addition, these antibodies primarily recognized linear epitopes in the variable domains and neutralized T-cell line-adapted (TCLA) virus isolates at significant dilutions,[121,122] and fail to neutralize primary HIV-1 isolates.[3] Furthermore, these non-conformational anti-gp120 specific antibodies are predominantly subtype-specific and, to some extent, isolate-specific. However, contrary to the earlier neutralization data, a recently published study indicated that using a modified neutralization assay with an extended incubation phase, these first generation of monomeric gp120 vaccines can induce antibodies capable of neutralizing primary isolates.[123] Thus for the fair comparisons of potential vaccine candidates there is an urgent need to develop standardized neutralization assays.

Another factor that has the potential to limit the protective effect of neutralizing antibodies is the emergence of neutralization escape mutants.[124] Richman and colleagues have shown that most patients with primary HIV infection (0–39 months) developed significant neutralizing antibody responses against the autologous virus; however, neutralizing responses against the laboratory-adapted isolates as well as heterologous primary isolates are slow to develop and often lower.[118] The appearance of neutralization-resistant mutants indicates that neutralizing antibody responses have a negative effect on the viral replication during HIV infection. Thus, it appears that the virus is under intense selection pressure to mutate neutralizing epitopes without compromising functional aspects of the Env, such as binding to the receptor and co-receptor to evade immune pressure. This adds additional constraints in development of an effective anti-HIV vaccine. Despite this selective pressure, virus replication proceeds unhindered in infected individuals. Thus, HIV Env can tolerate multiple mutations and most of these mutations are positioned in variable regions that contain several glycosylation sites. In general the number of glycosylation sites in Env glycoprotein is relatively conserved across subtypes and isolates, suggesting a pivotal role for carbohydrates in Env function and structural integrity. This extensive glycosylation provides the virus with some flexibility to modify neutralizing epitopes to evade the immune pressure simply by altering the glycosylation profile (adding, deleting or changing the position), without perturbing the secondary structure.[120]

Strategies to Design Immunogens that may Induce Neutralizing Antibodies of Protective Specificities by Vaccination

It is evident that gp120 contains neutralizing epitopes and that antibodies directed against these epitopes can protect against virus infection. During the course of natural infection, primary neutralizing antibody responses are induced in humans. However, the gp120 monomer has been relatively ineffective at eliciting these primary isolate neutralizing responses. Furthermore, a recently concluded phase III

efficacy trial using a gp120 monomeric vaccine demonstrated that there was no difference in the rate of infectivity between vaccine recipients and the placebo group.[125–127] The ability of this vaccine to raise functional antibody responses such as neutralizing Abs and ADCC has yet to be fully evaluated, but the information gained may help to guide further vaccine development. It will also be critical to understand the barriers in inducing antibodies that recognize similar epitopes to those recognized by 2F5, 4E10, Z13, F105, 2G12, and b12. Several strategies are being evaluated by various investigators to develop and evaluate novel Env immunogens to induce broadly neutralizing antibody responses by vaccination, including: (A) structural optimization targeting conserved neutralization epitopes, and (B) approaches to overcome genetic diversity for vaccine development.

A. *Structural Optimization to Target Conserved Neutralization Epitopes*

The main approaches toward structural optimization of Env for inducing broadly cross-reactive neutralizing antibodies involve the use of: (i) native trimers, (ii) triggered-Envs, (iii) rationally designed Env structures, and iv) mimotopes.

(i) *Native trimers as immunogen*

As mentioned previously, critical neutralizing epitopes are preserved and presented on the Env glycoprotein, but the virus has evolved ways to minimize the immunogenicity of Env, such as by extensive glycosylation of the molecule, to shield the conserved functional regions. Similarly, the receptor-binding site, a likely target for immune intervention, is buried within the molecule and is partially protected from immune recognition. The observation that most broadly reactive neutralizing antibodies isolated so far (i.e. IgG1b12, 2G12, and 2F5) have a stronger affinity for the native envelope (trimer) than for monomeric gp120 or gp41 provides[44,128,129] the impetus to evaluate trimeric Env glycoproteins for their ability to induce broadly cross-reactive, primary isolate neutralizing antibodies. Early immunogenicity

studies demonstrated that antibodies induced by oligomeric Env cross-reacted with HIV Env of other subtypes and neutralized both T cell-adapted and some primary isolates of HIV-1.[130] To use the trimeric Env as an immunogen to elicit functional antibody responses, the first and foremost challenge is to stabilize Env in the trimeric conformation without compromising the structural integrity of the molecule. Several approaches have been tried with reasonable success.[78,131–142] Yang *et al.* used GCN4-stabilized oligomers to demonstrate that antibodies induced by oligomeric gp140 were more effective at neutralizing heterologous primary isolates compared to antibodies elicited by the corresponding monomeric gp120 protein.[143] Earl *et al.* made similar observations in rhesus macaques using a trimeric Env protein from SIV.[135] Recently, we reported the purification and characterization of stable trimeric Env protein from US4 (Subtype B R5 isolates).[140] We also purified and characterized trimeric protein from SF162 and demonstrated that it elicited strong antibody responses in rabbits that neutralized two out of six heterologous subtype-B primary isolates.[144] To further enhance the immunogenicity of trimeric Env, we have introduced a partial deletion of the V2 loop and expressed, purified, and characterized this novel Env immunogen.[139] Using biophysical, biochemical, and immunological techniques, we demonstrated that the purified Env was a trimer, had nanomolar affinity to CD4, and had critical neutralizing epitopes exposed and preserved on the trimer.[139] Deletion of the V2 loop qualitatively altered the immunogenicity of the V3 and V1 loops and rendered the C5 region more immunogenic.[145] The V2 loop deletion in the context of a trimer elicited strong antibodies in rabbits that neutralized five out of six heterologous subtype-B primary isolates, and conferred partial protection upon pathogenic challenge with $SHIV_{SF162P4}$ to rhesus macaques immunized with this V2 loop deleted trimer in a DNA prime and protein boost strategy.[146,147] These animals were followed for more than three years without any sign of disease. In addition, we demonstrated a reasonable correlation between the presence of neutralizing antibodies at the time of challenge in the vaccinated animals and the levels of plasma viremia during acute infection.[148]

This protein is currently being evaluated in a human clinical trial (HVTN049). In a separate study, Quinnan and colleagues have evaluated the efficacy of oligomeric Env for inducing broadly neutralizing antibodies.[149] They primed a group of rhesus macaques with alphavirus replicon particles expressing the Env from R2 strain, and later these animals were boosted with the soluble oligomeric gp140. Concurrently, animals were also immunized with SIV gag/pol genes or no SIV genes to determine the additional protective benefit of inducing cell-mediated immune responses. The antibodies induced by alphavirus priming and oligomeric protein boost neutralized diverse primary isolates *in vitro*. The immunized animals were protected against the challenge infection with SHIVDH12 (Clone7). Furthermore, the protection was associated with neutralizing antibody titers of >1:80 or with cellular responses of 2000 spot-forming cells/10E6 PBMC. More recently, we have extended these studies for the subtype C as well. In a proof of concept study in both rabbits and rhesus macaques, we have demonstrated that HIV envelope DNA vaccines derived from the South African subtype-C TV1 strain can effectively prime for humoral responses against both subtype-B and subtype-C primary R5-tropic HIV-1 isolates in rabbits and rhesus macaques. Priming rabbits with DNA plasmids encoding TV1 gp140 modified in the second hypervariable loop (gp140TV1ΔV2), followed by boosting with oligomeric TV1ΔV2 proteins, elicited more potent HIV-neutralizing antibody responses than DNA vaccines encoding TV1 gp160 or the intact TV1 gp140. Boosting with oligomeric TV1ΔV2 envelope proteins in MF59 adjuvant also elicited higher titers of antibodies with homologous neutralizing activity against TV-1 and heterologous neutralizing activity against the subtype-B SF162 primary isolate. In addition, combining subtype-B and subtype-C V2-deleted immunogens resulted in increased cross-clade neutralizing activity. Importantly, subtype-B-neutralizing antibody responses were also observed after immunizing rhesus macaques with subtype-C TV1ΔV2 immunogens. Our results demonstrate that the modified subtype-C gp140TV1ΔV2 immunogen described herein elicits broad, subtype-B and subtype-C neutralizing antibodies.[150]

(ii) *Triggered Env: targeting conformational epitopes*

Conformational epitopes in Env induced upon interaction with CD4 are attractive targets for inducing neutralizing antibodies. It is well documented that upon binding to CD4, HIV Env protein undergoes significant conformational changes,[151,152] which are critical for membrane fusion and viral entry in the target cells.[152] Furthermore, mAbs such as 17b, 48d, 23e, 49e, 21c, E51, CG10, and X5 are better able to recognize specific epitopes on gp120 after it interacts with soluble CD4 (triggered Env).[46,48,49,153,154] Attempts have been made by various groups to evaluate such triggered Env (gp120-CD4 complexes) as potential vaccine candidates.[47,155] It was demonstrated that gp120-CD4 complexes can induce broadly neutralizing antibody responses, but they need to be stabilized using cross-linking reagents.[155] In a proof of concept study, we have demonstrated that triggered Env (gp120-CD4 complexes) induced strong immune responses against both gp120 and CD4[145] and these antibodies neutralized both a primary isolate (SF162) as well as a T-cell adapted isolate (SF2).[156] To elucidate the contribution of anti-Env antibodies to neutralization, we affinity purified the anti-Env antibodies from anti-CD4 antibodies and showed that these antibodies neutralized two subtype-B and one subtype-C primary HIV-1 isolates that were tested.[156] Since gp120 SF2 alone was not able to induce primary isolate neutralizing antibodies,[144] these preliminary data suggested that certain neutralizing epitopes may be better exposed on triggered Env compared to untriggered gp120.

These studies illustrate the potential utility of the approach; however, the use of full-length CD4 raises the potential for induction of autoimmune responses. One strategy to overcome this problem is to identify the minimal binding domain of CD4. The recently reported crystallographic structure of gp120, in complex with CD4 and the Fab portion of mAb 17b,[157] has demonstrated that a large surface (742 Å) of the domain D1 of CD4 binds to a large depression (800 Å) on gp120. This CD4 interface is composed of 22 residues, contributing to gp120 binding with mixed hydrophobic, electrostatic, and H-bonding interactions. The large size and complexity of this

interface makes the reproduction of such a functional domain into a small non-immunogenic molecule a challenge, and highlights the difficulty in developing high-affinity CD4 mimics. However, despite of the large number of residues present in gp120-CD4 interaction, studies on hormone-receptor systems have shown that few residues might dominate the binding energy at protein-protein interface.[158] Thus, the design of a minimal CD4 mimics may be possible.

The transfer of functional sites to small proteins acting as structural scaffolds has been proposed as a useful strategy to reproduce the structure and function of the target protein in small molecular systems.[159,160] This approach has led to the discovery of scorpion toxin Scyllatoxin fragment as an effective mimic of CD4. A mini-protein, CD4M3, was chemically synthesized, folded efficiently, and presented a circular dichroism spectrum similar to that of native Scyllatoxin. In competitive ELISA, CD4M33 was able to specifically bind gp120 at an IC50 of 40 μM, which is four orders of magnitude higher than that of sCD4. This strategy has been recently applied to the engineering of a mini-protein that mimics the core of the CD4 protein surface that interacts with the gp120 envelope glycoprotein of HIV-1 and, hence, inhibits virus attachment to cells and infection.[161]

The biological performance of this mini-protein was improved using "rational" structure design. In total, five substitutions were introduced (Gln20Ala, Thr25Ala, Leu18Lys, Ser9Arg, and Pro28) and the resulting mini-protein (CD4M9) bound to gp120 at 400 nM, induced conformational changes in Env, and inhibited infection of CD4-expressing cells by different virus isolates. So far, an improved CD4M33 has an affinity for gp120 similar to CD4,[162] and induces a conformational change in gp120 similar to that induced by sCD4X.[162] In an earlier study, Fouts and colleagues have shown that covalently cross-linked complexes of CD4 and HIV Env IIIB induced antibodies that neutralized a wide range of primary isolates.[163] Ig with neutralizing activity was recovered by affinity chromatography using Env/CD4M9 single-chain polypeptide. More recently, using CD4M9, an earlier version of the CD4 mini-protein developed by Vita and colleagues, Fouts *et al.* prepared single-chain constructs of SCBal/M9 and performed immunogenicity studies in rabbits.[164] They showed

that antibodies induced by Env-CD4M9 cross-linked complexes neutralized a broad range of primary isolates. These results demonstrate that: i) a significant portion of gp120 binding surface of CD4 can be reproduced in a mini-protein system, ii) an engineered CD4 mini-protein contains sufficient CD4 structural elements to induce gp120 conformational changes, and iii) these surrogate molecules may be useful in making stable complexes with envelope protein to expose envelope epitopes for the induction of neutralizing antibodies.

In another approach, Guo *et al.* described a small synthetic molecule (termed BMS 378806) that inhibited the interaction of gp120 with cellular CD4 and prevented viral entry.[165] This compound bound to gp120 at a stoichiometry of approximately 1:1 with an affinity similar to that of CD4. Therefore, molecules such as this one should be evaluated for their ability to induce conformational change in Env.

(iii) *Rational design of Env immunogen*

A crystal structure of HIV-1 gp120 core in complex with two N-terminal domains of CD4 and a Fab fragment of mAb 17b that binds CD4-induced co-receptor-binding site was solved by Kwong,[166] and is known as liganded structure. The structure was determined using the gp120 that lacked the V1/V2 and V3 loops as well as 52 and 19 residues from the N- and C-termini. The structure provided some basic, but very important, information about the organization of gp120. The protein is comprised of two domains, the inner and the outer, which are bridged by a mini-domain called a "bridging sheet", a four-stranded, anti-parallel (β-sheet) (Fig. 7). The inner domain is buried internally in the gp120/gp41 trimeric complex; therefore, as expected, antibodies against this region do not neutralize the virus. The outer domain is extensively glycosylated and does not elicit antibody response, and thus it is termed as immunologically silent. The receptor binding site (i.e. CD4 binding pocket) is located in a region that borders the outer and inner domains and the bridging sheet. More recently, crystal structures of HIV-1 gp120 with intact V3 loop have been solved.[167] It appears that V3 loop forms a rod-like structure that protrudes out from the gp120 core. This is not surprising

considering its highly immunogenic properties. The crystal structure of SIV gp120 reveals a number of important facts. **First**, its overall structure is very similar to that of HIV-1 gp120, especially the outer domain of the protein.[76] **Second**, neither the CD4 nor the co-receptor binding site (BS) is properly formed in the unliganded state of the Env. In the unliganded form, the two double-stranded anti-parallel sheets are well separated and many residues that make up the CD4 BS are either masked or in different configuration compared to after binding to CD4. **Third**, there is large displacement and rotation of the inner domain and bridging sheet components upon CD4 binding: (i) the three-strand anti-parallel sheet of the inner domain rotates by 30°, (ii) the four-turn α-helix 1 shifts away from the outer domain, and (iii) the tip of the V1 $\alpha/2$ stem moves by over 40Å. These events result in the formation of the bridging sheet. At present, the structure of the V1/V2 loop is not known. However, it is thought to partially mask the bridging sheet. Deeper insights into the native (i.e. un-truncated) structures of gp120 could facilitate envelope-based vaccine design. Now the efforts are being made by different groups including ours to rationally design novel immunogens that may induce broadly neutralizing antibody responses. So far, the focus has been to optimize engineered Env structure for inducing potent antibody responses against conserved functional epitopes by structure-based targeted deglycosylation, by hyper-glycosylation, or by introducing mutations/deletions in the bridging sheet.

Structure-based Targeted Deglycosylation

The extensive glycosylation of Env is likely to be involved in immune evasion. Based on crystal structure it seems that the outer domain of gp120 is heavily glycosylated, as shown by Wyatt and colleagues[77] (Fig. 8). Despite being most exposed to antibodies, this region of Env is known as "silent face" because it seldom induces neutralizing antibody responses. Thus, the sugar moieties may be shielding critical neutralization epitopes. Malenbaum *et al.* have demonstrated that the removal of glycosylation at position 301 resulted in an increased neutralizing sensitivity of HIV-1 to CD4 BS antibodies.[168] Furthermore,

mutant virus lacking 301 glycan also demonstrated sensitivity to CD4i antibodies. In another study, elimination of N-linked glycosylation in the V1 and V2 loops of pathogenic SIV mac239 rendered the virus more sensitive to host antibody responses.[169] It is apparent from the crystal structure that sugar moieties lie proximal to, but not within, the receptor binding site. Koch and colleagues have performed structure-based deglycosylation of four sites flanking the receptor-binding region (i.e. 197, 276, 301, and 386). Removal of a single glycosylation site at the base of V3 loop (i.e. aa 301) has rendered the mutant virus more sensitive to antibody-mediated neutralization by anti-CD4 BS antibodies. Furthermore, deletion of all 4 glycosylation sites has made the resultant virus sensitive to neutralization by CD4i antibodies. In an other study, McCaffrey *et al.*[170] have demonstrated that removal of sugar moieties at positions 293, 299 329, and also 438 and 454 made the SF162 virus more sensitive to neutralization by CD4BS mAbs and also to V3 loop specific neutralizing mAbs. In addition, the deglycosylation of V3 loop (293, 299, and 329) also made the mutant virus more sensitive to mAbs specific to CD4i epitope. Furthermore, the same group has also demonstrated that deglycosylation at positions 293 (V3 loop), 438 (C4 region), and 454 (V5 loop) also rendered the resultant virus more susceptible to the anti-gp41 mAb 2F5. These studies suggest that carbohydrates at these positions protect the CD4 binding site, V3 loop, co-receptor binding site and also gp41 epitope.[170] Further studies will be needed to evaluate the efficacy of this approach in exposing critical neutralizing epitopes in Env immunogens.

Hyper-glycosylation for Focusing the Immune Response Toward Neutralizing Epitopes Recognized by b12

Burton and colleagues have taken the opposite approach of incorporating additional carbohydrate residues. They have demonstrated that four alanine substitutions on the perimeter of the Phe43 cavity of gp120 reduced the binding of weakly neutralizing CD4 BS antibodies

to gp120,[171] while increasing the binding of a potent broadly neutralizing antibody b12. In further studies, they focused on the reduction of binding of a wide range of non-neutralizing antibodies by incorporating seven more glycosylation sites in addition to the four alanine mutations.[172] It was interesting to note that these hyperglycosylated Env proteins were not recognized by non-neutralizing antibodies directed toward CD4 BS (such as b3, b6, CD4-IgG2, 15e, F91, and F105); however, binding of the neutralizing antibody b12 remained, albeit at lower affinity, largely unaffected. Furthermore, hyperglycosylation affected the exposure of conformational epitopes recognized by mAbs 17b, 48d, and X5. It remains to be determined whether these modified molecules can alter the immunogenicity of Env and, in particular, if the increased b12 reactivity can be correlated with enhanced neutralizing activity.

Directing the Immune Response Toward the CD4 Binding Site by Introducing Deletions in Bridging Sheet

It has been difficult to induce antibody responses directed against the CD4 BS by vaccination with gp120. From gp120 structure analysis, it is apparent that Env is folded into inner and outer domains (Fig. 7). The inner domain (with respect to the N and C terminus) comprises two helices, a small five-stranded β sheet sandwiched at its proximal end, and a projection at the distal end from which the V1/V2 loop emanates. The outer domain is a stalked double barrel that lies alongside the inner domain, such that the outer barrel and inner bundle axes are approximately parallel. The gp120 inner and outer domains are attached by a bridging sheet that further limits the accessibility of the CD4 binding pocket. This bridging sheet is composed of four anti-parallel β strands (namely β-2, β-3, β-20, and β-21) (Fig. 8). The CD4 binding site is buried deep between the inner and outer domains of this molecule and, hence, possibly not accessible to antibodies. Structural data suggest that the V2 loop may fold over the bridging sheet and, due to its three-dimensional position, the bridging sheet is

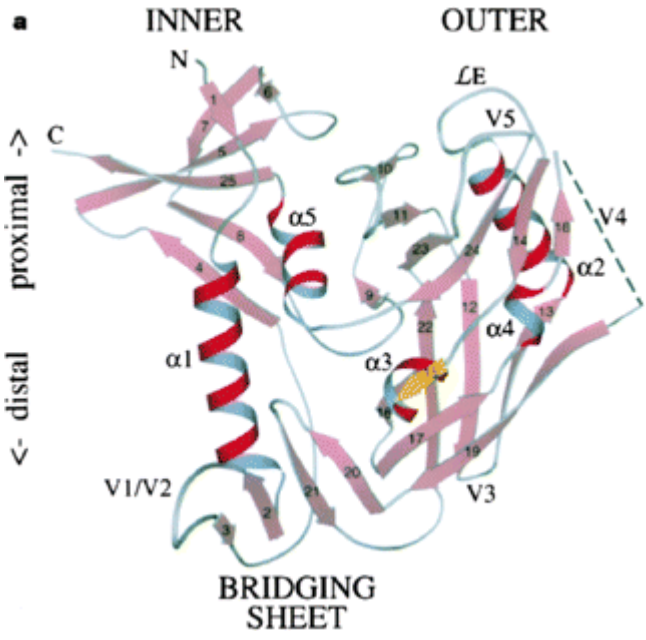


Fig. 7. gp120 core structure. Schematic representation of the tertiary structure of the HIV-1_{HXB-2} Env gp120 polypeptide highlighting the bridging sheet that connects the outer and inner domains as determined by the crystallographic studies.[157]

believed to mask the elements involved in CD4 binding and co-receptor binding. It has been shown that the deletion of V1 or V1V2 loops does not abrogate the functional activities of the envelope glycoprotein because the recombinant V1 or V1V2 deleted viruses are fully replication-competent in human PBMC.[173] Furthermore, these deletions in the context of Env protein also did not alter its ability to bind sCD4.[139,174] This suggests that these modifications alone preserve the integrity of important conserved epitopes on Env. Furthermore, deletion of V1 and V2 loops has made the virus more susceptible to antibody-mediated neutralization; however, no improvement was observed in terms of directing the immune response to the CD4 binding site.[139] Therefore, we introduced additional deletions in the bridging sheet in an attempt to sufficiently expose the CD4 binding

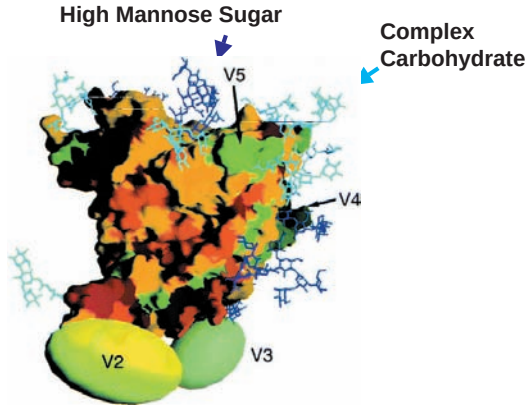


Fig. 8. The molecular surface of the gp120 core, including the modeled N-terminal residues, V4 loop, and carbohydrate structures, are shown as presented by Wyatt *et al.*[77] The outer domain of gp120 is heavily decorated with complex carbohydrates (blue) and terminal mannose sugars (dark blue). The molecular surface of gp120 is color-coordinated to demonstrate the variability in the Env: red indicates residues conserved among all primate immunodeficiency viruses; orange indicates residues conserved in all HIV isolates; yellow indicates some degree of variability; and green indicates significant variability among all HIV isolates.

site without destroying the proper folding and conformation of Env. As mentioned above, there are four β strands: β-2 strand corresponds to approximately amino acid residues Cys-119 to Thr-123, β-3 corresponds to approximately Ser199 to Ile-201, β-20 extends from amino acid residues 422-Gln to 426-Met, and β-21 extends from amino acid residues 431-Gly to 435-Tyr relative to HXB-2. The V1 and V2 loops emanate from the first pair of β strands (Cys-126 to Cys-196) and a small loop extrudes from the second set of β strands. This small loop extends from amino acid Trp-427 to Val-430. The H-bonds between β-2 and β-21 are the only connections between domains of the lower half of the protein (joining helix α-1 to the CD4 BS). Based on these structural features, we have designed a series of deletions within the small loop of the bridging sheet to further expose the CD4 binding site. In addition, we are evaluating deletions in the large loop (i.e. V1 and V2 loops) either alone or in conjunction with

deletions in small loop to enhance the exposure of conserved neutralization epitopes that are shielded from the immune system (unpublished observations).

Structural Information on gp41

Four crystal structures of HIV/SIV gp41 cores are available; all of these structures are similar and represent a post-fusion conformation of gp41.[175] It seems that in prefusion state, in the native and untriggered conformation, the surface of gp41 is shielded by gp120, and thus is not accessible for antibody recognition. Upon binding of CD4 and co-receptor, both gp120 and gp41 undergo extensive conformational changes that expose the gp41 in an extended fusion intermediate conformation to position the fusion peptides for insertion into the host membrane, during which gp41 is perhaps most vulnerable from host antibody binding. In a proof of concept study, LaCasse and colleagues have shown that it is possible to target the fusion intermediate for vaccine application.[176] In the study they immunized the animals with fusion intermediates, and demonstrated that post immunization sera neutralized 24 out of 25 primary isolated tested. Furthermore, it has been shown that the peptides derived from the sequences of HR1 and HR2 blocked the membrane fusion process of HIV.[76,177] Most likely, these peptides bind to the gp41 fusion intermediates during the conformational changes, and therefore prevent the membrane fusion and viral internalization. Now the most important question is how to exploit the potential of fusion intermediates for vaccine development. The most important constraint is exploiting these fusion intermediates for vaccine application is that the structure of gp41 in fusion intermediate conformation is not known. Therefore, structural information of a longer or full-length gp41 protein will be helpful to understand the influence on overall gp41 structure and function of the segments that were omitted in the present crystal structures of gp41 core. In addition, detailed structure studies of gp41 variants in conjunction with functional characterization of antigenicity of the forms can drive a better understanding of the structural epitopes to productively drive antigen development.

Current Structural Information that may be used for Designing new Immunogens

While the efforts were directed to solve the crystal structure of Env trimer, and to address the technical challenges associated with it, two groups have applied a low-resolution technique (i.e. Cryo-EM) to circumvent the problems associated with crystallization and to provide the structural information on the Env trimer.[178,179] Zulleti *et al.* have two best-fitted Cryo-EM three-dimensional desnsities, and both of these structures suggest that: i) all the glycans are pointing outward from the surface of the trimer, and solvent exposed, which is consistent with the glycan shield model proposed by different investigators to protect the virus against the neutralizing antibodies[120,180]; ii) the variable regions V4 and V5 are positioned on the top surface of the complex, an orientation that would facilitate their recognition by antibodies[181]; iii) the V1V2 loop extends outward, implying good solvent exposure, and in addition it is possible that the interaction of V1V2 loop and the co-receptor binding site in the monomer may limit the access of the co-receptor binding site to antibodies; iv) the CD4 binding surface is exposed on the outer edge of each gp120 protomer, and is oriented such that access to membrane-associated CD4 on the target cell is possible; and v) the two models differ significantly with regard to the orientation and exposure of V3 loop, and the orientation of co-receptor binding sites. In the **first model**, the ***V3 base points toward the trimer interface*** running roughly parallel to the 3-fold symmetry axis, being partially exposed in the cavities between the lobes. In this trimer model, the V3 loop would be substantially masked by packing into the trimer axis, potentially reinforced by inter-V3 bonding. In the **second model**, the ***V3 loop is oriented outward in solvent exposed phase***. In addition, V3 seems to be highly flexible in the CD4-unbound trimer. After Env binds to CD4 it induces a conformational change in Env that increases V3 exposure. This is supported experimentally by the increased accessibility of the V3 loop in the Env trimer to antibody binding and enzymic proteolysis subsequent to CD4 engagement.[182,183] It is also consistent with the highly exposed nature of the V3 loop in the structure of the

V3-containing gp120–CD4–Fab complex.[167] The increased exposure of V3 in the CD4-bound conformation could be explained by a rearrangement of the trimer in the first model, or by a conformational change involving the V3 loop in the second model.

The conserved **co-receptor binding site**, comprising the bridging sheet[76] and associated regions, is thought to be largely inaccessible in the trimer. Experimental data support this concept, since most CD4-induced (CD4i) surface-specific mAbs cannot access their epitopes on the CD4-unligated Env trimer, as evidenced by weak or absent neutralization. It is possible that the V3 loop may be involved in protecting the co-receptor binding sites from antibody recognition and binding.

A trimer model has been proposed previously in which the inner domain of gp120 points toward the gp120-gp41 interface and the outer domain extends outward.[76] Such a model is inconsistent with the orientation of the major axis of the ellipsoidal gp120 density within the density corresponding to gp120 in our structure. In this previous model, the V3 loop is exposed to solvent and the co-receptor binding site is buried at the trimer interface. The two fittings proposed by Zanetti *et al.* clearly suggest that either the V3 loop is also pointing toward the trimer interface, or the co-receptor binding site is on the outside of the trimer, and protected from solvent by V3.[178] It is likely that the Env complex exhibits some conformational flexibility.

In the context of trimeric Env, the gp41 stem appears as a compact structure with no obvious separation between the three monomers.[178] The membrane-proximal region of gp41 is characterized by a highly conserved hydrophobic region, which is thought to mediate trimer self-assembly.[184] Furthermore, the density corresponding to the gp41 stem region is shorter and slightly wider than the post-activation coiled-coil conformation,[185] in agreement with the hypothesis that a dramatic conformational change is required for gp41 to extend toward, and insert into, the target cell membrane. The shape of the complex suggests a new model for the Env trimer organization. The volumes are consistent with the assignment of the globular domains to gp120 and the stem to gp41. The gp120

protomers appear to fold over gp41 rather than depart radially from it, contacting each other at the top of the spike. The interaction between gp120 monomers at this contact is likely to be weak, to allow the disassembly of Env ectodomain and extrusion of gp41 for insertion into the host membrane for the membrane fusion to take place.

The two models proposed shed light on several important antigenic and mechanistic features of the Env trimer. First, they provide models consistent with the concept that glycans and immunodominant variable loops are positioned on exposed gp120 surfaces to damp the neutralizing antibody response. Second, they confirm that only a limited gp41 surface is exposed for antibody binding, as has been proposed previously.[92] Third, they provide two possible descriptions of the position of the V3 loop and the co-receptor binding.

B. *Approaches to Overcome Genetic Diversity of HIV Env for Vaccine Development*

HIV is one of the most genetically diverse viral pathogens studied to date. The uneven distribution of the various clades across the globe presents one of the most challenging aspects to HIV vaccine design, as it is unknown whether it will be possible to make an HIV vaccine to cover all clades or whether a tailor-made vaccine based on the most prevalent strain for a given region will be needed. To address the problem of genetic diversity in HIV multivalent vaccine formulations and vaccines based on consensus and ancestral sequences are being evaluated.

Multivalent vaccine approach

The objective of this approach is to use multiple clade-specific immunogens in a vaccine to increase the breadth of the immune responses without compromising the potency against one or more clades. Various groups have recently demonstrated the potential utility of this approach. For example, Shan Lu and colleagues demonstrated that by including monomeric Envs from different clades, improved breadth of neutralizing antibodies against different subtypes

was achieved.[186] Similar observations were made with a DNA prime and rAdeno virus boost regimen by Letvin and colleagues.[187] In a proof of concept study, we immunized a group of rabbits with HIV-1 Envs derived from subtype-B and subtype-C isolates either alone or in combination in a DNA prime and protein boost regimen. All of the animals immunized either alone or with the bivalent B and C vaccine induced comparable levels of antibody responses against subtype-B and subtype-C envelope proteins in ELISA. The antibodies induced in the bivalent group were able to neutralize both homologous subtype-B and subtype-C isolates as well as other B and C strains.[150]

Acknowledgments

We would like to thank Dr Rino Rappuoli, Vice President, Novartis Vaccines and Diagnostics, Inc. for his encouragement and support of the HIV vaccine development project, and Dr Susan Barnett (HIV Project Leader) and Dr Jeffrey Ulmer (Head, Immunology and Cell Biology) for their valuable contributions. We would also like to thank K.C. Egan for their expert editorial help and Mike Peterson for administrative help and support. HIV vaccine work is supported by NIAID-NIH contracts and grants (AI-95367; I-AI-05396 and 5 PO1 AI48225-03), as well as the Discovery Fund.

References

1. UNAIDS/WHO. (2004) *2004 Report on the global AIDS epidemic.* Online.
2. Mascola JR, Mathieson BJ, Zack PM, *et al.* (1993) Summary report: workshop on the potential risks of antibody-dependent enhancement in human HIV vaccine trials. *AIDS Res Hum Retroviruses* **9**: 1175–1184.
3. Mascola JR, Snyder SW, Weislow OS, *et al.* (1996) Immunization with envelope subunit vaccine products elicits neutralizing antibodies against laboratory-adapted but not primary isolates of human immunodeficiency virus type 1. The National Institute of Allergy and Infectious Diseases AIDS Vaccine Evaluation Group. *J Infect Dis* **173**: 340–348.

4. Barouch DH, Kunstman J, Glowczwskie J, *et al.* (2003) Viral escape from dominant simian immunodeficiency virus epitope-specific cytotoxic T lymphocytes in DNA-vaccinated rhesus monkeys. *J Virol* **77**: 7367–7375.
5. Barouch DH, Kunstman J, Kuroda MJ, *et al.* (2002) Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. *Nature* **415**: 335–339.
6. Barouch DH, Letvin NL. (2002) Viral evolution and challenges in the development of HIV vaccines. *Vaccine* **20**: A66–68.
7. Vogel TU, Reynolds MR, Fuller DH, *et al.* (2003) Multispecific vaccine-induced mucosal cytotoxic T lymphocytes reduce acute-phase viral replication but fail in long-term control of simian immunodeficiency virus SIVmac239. *J Virol* **77**: 13348–13360.
8. Mascola JR, Lewis MG, VanCott TC, *et al.* (2003) Defining the protective antibody response for HIV-1 cellular immunity elicited by human immunodeficiency virus type 1/ simian immunodeficiency virus DNA vaccination does not augment the sterile protection afforded by passive infusion of neutralizing antibodies. *Curr Mol Med* **3**: 209–216.
9. Hilleman MR. (2001) Overview of the pathogenesis, prophylaxis and therapeusis of viral hepatitis B, with focus on reduction to practical applications. *Vaccine* **19**: 1837–1848.
10. Robbins JB, Schneerson R, Szu SC. (1995) Perspective: hypothesis: serum IgG antibody is sufficient to confer protection against infectious diseases by inactivating the inoculum. *J Infect Dis* **171**: 1387–1398.
11. Keller MA, Stiehm ER. (2000) Passive immunity in prevention and treatment of infectious diseases. *Clin Microbiol Rev* **13**: 602–614.
12. Klasse PJ, Sattentau QJ. (2001) Mechanisms of virus neutralization by antibody. *Curr Top Microbiol Immunol* **260**: 87–108.
13. Yang X, Kurteva S, Ren X, *et al.* (2005) Stoichiometry of envelope glycoprotein trimers in the entry of human immunodeficiency virus type 1. *J Virol* **79**: 12132–12147.
14. Binley JM, Cayanan CS, Wiley C, *et al.* (2003) Redox-triggered infection by disulfide-shackled human immunodeficiency virus type 1 pseudovirions. *J Virol* **77**: 5678–5684.
15. Chan DC, Kim PS. (1998) HIV entry and its inhibition. *Cell* **93**: 681–684.
16. Golding H, Zaitseva M, de Rosny E, *et al.* (2002) Dissection of human immunodeficiency virus type 1 entry with neutralizing antibodies to gp41 fusion intermediates. *J Virol* **76**: 6780–6790.

17. Gorny MK, Zolla-Pazner S. (2000) Recognition by human monoclonal antibodies of free and complexed peptides representing the prefusogenic and fusogenic forms of human immunodeficiency virus type 1 gp41. *J Virol* **74**: 6186–6192.
18. Hansen JE, Clausen H, Nielsen C, *et al.* (1990) Inhibition of human immunodeficiency virus (HIV) infection *in vitro* by anticarbohydrate monoclonal antibodies: peripheral glycosylation of HIV envelope glycoprotein gp120 may be a target for virus neutralization. *J Virol* **64**: 2833–2840.
19. Scanlan CN, Pantophlet R, Wormald MR, *et al.* (2002) The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of alpha1 → 2 mannose residues on the outer face of gp120. *J Virol* **76**: 7306–7321.
20. Ugolini S, Mondor I, Parren PW, *et al.* (1997) Inhibition of virus attachment to $CD4^+$ target cells is a major mechanism of T cell line-adapted HIV-1 neutralization. *J Exp Med* **186**: 1287–1298.
21. Liao HX, Alam SM, Mascola JR, *et al.* (2004) Immunogenicity of constrained monoclonal antibody A32-human immunodeficiency virus (HIV) Env gp120 complexes compared to that of recombinant HIV type 1 gp120 envelope glycoproteins. *J Virol* **78**: 5270–5278.
22. Dimmock NJ. (1993) Neutralization of animal viruses. *Curr Top Microbiol Immunol* **183**: 1–149.
23. Ahmad A, Menezes J. (1996) Antibody-dependent cellular cytotoxicity in HIV infections. *Faseb J* **10**: 258–266.
24. Ahmad R, Sindhu ST, Toma E, *et al.* (2001) Evidence for a correlation between antibody-dependent cellular cytotoxicity-mediating anti-HIV-1 antibodies and prognostic predictors of HIV infection. *J Clin Immunol* **21**: 227–233.
25. Gomez-Roman VR, Cao C, Bai Y, *et al.* (2002) Phage-displayed mimotopes recognizing a biologically active anti-HIV-1 gp120 murine monoclonal antibody. *J Acquir Immune Defic Syndr* **31**: 147–153.
26. Spear GT, Takefman DM, Sharpe S, *et al.* (1994) Antibodies to the HIV-1 V3 loop in serum from infected persons contribute a major proportion of immune effector functions including complement activation, antibody binding, and neutralization. *Virology* **204**: 609–615.
27. Alsmadi O, Tilley SA. (1998) Antibody-dependent cellular cytotoxicity directed against cells expressing human immunodeficiency virus type 1 envelope of primary or laboratory-adapted strains by human and chimpanzee monoclonal antibodies of different epitope specificities. *J Virol* **72**: 286–293.

28. Sawyer LA, Katzenstein DA, Hendry RM, *et al.* (1990) Possible beneficial effects of neutralizing antibodies and antibody-dependent, cell-mediated cytotoxicity in human immunodeficiency virus infection. *AIDS Res Hum Retroviruses* **6**: 341–356.
29. Forthal DN, Landucci G, Daar ES. (2001) Antibody from patients with acute human immunodeficiency virus (HIV) infection inhibits primary strains of HIV type 1 in the presence of natural-killer effector cells. *J Virol* **75**: 6953–6961.
30. Ljunggren K, Moschese V, Broliden PA, *et al.* (1990) Antibodies mediating cellular cytotoxicity and neutralization correlate with a better clinical stage in children born to human immunodeficiency virus-infected mothers. *J Infect Dis* **161**: 198–202.
31. Battle-Miller K, Eby CA, Landay AL, *et al.* (2002) Antibody-dependent cell-mediated cytotoxicity in cervical lavage fluids of human immunodeficiency virus type 1-infected women. *J Infect Dis* **185**: 439–447 Epub 2002 Jan 31.
32. Gomez-Roman VR, Patterson LJ, Venzon D, *et al.* (2005) Vaccine-elicited antibodies mediate antibody-dependent cellular cytotoxicity correlated with significantly reduced acute viremia in rhesus macaques challenged with SIVmac251. *J Immunol* **174**: 2185–2189.
33. Richard N, Iacampo S, Cochrane A. (1994) HIV-1 Rev is capable of shuttling between the nucleus and cytoplasm. *Virology* **204**: 123–131.
34. Moore JP, Doms RW. (2003) The entry of entry inhibitors: a fusion of science and medicine. *Proc Natl Acad Sci USA* **100**: 10598–10602 Epub 2003 Sep 05.
35. Karwowska S, Gorny MK, Buchbinder A, *et al.* (1992) Production of human monoclonal antibodies specific for conformational and linear non-V3 epitopes of gp120. *AIDS Res Hum Retroviruses* **8**: 1099–1106.
36. Pinter A, Honnen WJ, Racho ME, Tilley SA. (1993) A potent, neutralizing human monoclonal antibody against a unique epitope overlapping the CD4-binding site of HIV-1 gp120 that is broadly conserved across North American and African virus isolates. *AIDS Res Hum Retroviruses* **9**: 985–996.
37. Poignard P, Fouts T, Naniche D, *et al.* (1996) Neutralizing antibodies to human immunodeficiency virus type-1 gp120 induce envelope glycoprotein subunit dissociation. *J Exp Med* **183**: 473–484.
38. Barbas CF, 3rd, Bjorling E, Chiodi F, *et al.* (1992) Recombinant human Fab fragments neutralize human type 1 immunodeficiency virus in vitro. *Proc Natl Acad Sci USA* **89**: 9339–9343.

39. Binley JM, Wrin T, Korber B, *et al.* (2004) Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. *J Virol* **78**: 13232–13252.
40. Kessler JA, 2nd, McKenna PM, Emini EA, *et al.* (1997) Recombinant human monoclonal antibody IgG1b12 neutralizes diverse human immunodeficiency virus type 1 primary isolates. *AIDS Res Hum Retroviruses* **13**: 575–582.
41. Trkola A, Pomales AB, Yuan H, *et al.* (1995) Cross-clade neutralization of primary isolates of human immunodeficiency virus type 1 by human monoclonal antibodies and tetrameric CD4-IgG. *J Virol* **69**: 6609–6617.
42. D'Souza MP, Milman G, Bradac JA, *et al.* (1995) Neutralization of primary HIV-1 isolates by anti-envelope monoclonal antibodies. *AIDS* **9**: 867–874.
43. Hioe CE, Xu S, Chigurupati P, *et al.* (1997) Neutralization of HIV-1 primary isolates by polyclonal and monoclonal human antibodies. *Int Immunol* **9**: 1281–1290.
44. Roben P, Moore JP, Thali M, *et al.* (1994) Recognition properties of a panel of human recombinant Fab fragments to the CD4 binding site of gp120 that show differing abilities to neutralize human immunodeficiency virus type 1. *J Virol* **68**: 4821–4828.
45. Ollman Saphire E, Parren PW, Barbas CF, *et al.* (2001) Crystallization and preliminary structure determination of an intact human immunoglobulin, b12: an antibody that broadly neutralizes primary isolates of HIV-1. *Acta Crystallogr D Biol Crystallogr* **57**: 168–171.
46. Thali M, Moore JP, Furman C, *et al.* (1993) Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J Virol* **67**: 3978–3988.
47. Gershoni JM, Denisova G, Raviv D, *et al.* (1993) HIV binding to its receptor creates specific epitopes for the CD4/gp120 complex. *Faseb J* **7**: 1185–1187.
48. Moulard M, Phogat SK, Shu Y, *et al.* (2002) Broadly cross-reactive HIV-1-neutralizing human monoclonal Fab selected for binding to gp120-CD4-CCR5 complexes. *Proc Natl Acad Sci USA* **99**: 6913–6918.
49. Xiang SH, Doka N, Choudhary RK, *et al.* (2002) Characterization of CD4-induced epitopes on the HIV type 1 gp120 envelope glycoprotein recognized by neutralizing human monoclonal antibodies. *AIDS Res Hum Retroviruses* **18**: 1207–1217.

50. Alkhatib G, Berger EA, Murphy PM, Pease JE. (1997) Determinants of HIV-1 coreceptor function on CC chemokine receptor 3. Importance of both extracellular and transmembrane/cytoplasmic regions. *J Biol Chem* **272**: 20420–20426.
51. Deng H, Liu R, Ellmeier W, *et al.* (1996) Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**: 661–666.
52. Feng Y, Broder CC, Kennedy PE, Berger EA. (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**: 872–877.
53. Dey B, Del Castillo CS, Berger EA. (2003) Neutralization of human immunodeficiency virus type 1 by sCD4-17b, a single-chain chimeric protein, based on sequential interaction of gp120 with CD4 and coreceptor. *J Virol* **77**: 2859–2865.
54. Labrijn AF, Poignard P, Raja A, *et al.* (2003) Access of antibody molecules to the conserved coreceptor binding site on glycoprotein gp120 is sterically restricted on primary human immunodeficiency virus type 1. *J Virol* **77**: 10557–10565.
55. Sodroski J. (2002) *Retroviruses of Human AIDS and Related Animal Diseases*. Elsevier, Paris.
56. Sullivan N, Sun Y, Binley J, *et al.* (1998) Determinants of human immunodeficiency virus type 1 envelope glycoprotein activation by soluble CD4 and monoclonal antibodies. *J Virol* **72**: 6332–6338.
57. Goudsmit J, Debouck C, Meloen RH, *et al.* (1988) Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. *Proc Natl Acad Sci USA* **85**: 4478–4482.
58. Javaherian K, Langlois AJ, McDanal C, *et al.* (1989) Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. *Proc Natl Acad Sci USA* **86**: 6768–6772.
59. Putney SD, Matthews TJ, Robey WG, *et al.* (1986) HTLV-III/LAV-neutralizing antibodies to an E. coli-produced fragment of the virus envelope. *Science* **234**: 1392–1395.
60. Gorny MK, VanCott TC, Hioe C, *et al.* (1997) Human monoclonal antibodies to the V3 loop of HIV-1 with intra- and interclade cross-reactivity. *J Immunol* **159**: 5114–5122.
61. Gorny MK, Xu JY, Karwowska S, *et al.* (1993) Repertoire of neutralizing human monoclonal antibodies specific for the V3 domain of HIV-1 gp120. *J Immunol* **150**: 635–643.

62. Conley AJ, Gorny MK, Kessler JA, 2nd, *et al.* (1994) Neutralization of primary human immunodeficiency virus type 1 isolates by the broadly reactive anti-V3 monoclonal antibody, 447-52D. *J Virol* **68**: 6994–7000.
63. Krachmarov CP, Kayman SC, Honnen WJ, *et al.* (2001) V3-specific polyclonal antibodies affinity purified from sera of infected humans effectively neutralize primary isolates of human immunodeficiency virus type 1. *AIDS Res Hum Retroviruses* **17**: 1737–1748.
64. Poignard P, Moulard M, Golez E, *et al.* (2003) Heterogeneity of envelope molecules expressed on primary human immunodeficiency virus type 1 particles as probed by the binding of neutralizing and nonneutralizing antibodies. *J Virol* **77**: 353–365.
65. Gorny MK, Williams C, Volsky B, *et al.* (2002) Human monoclonal antibodies specific for conformation-sensitive epitopes of V3 neutralize human immunodeficiency virus type 1 primary isolates from various clades. *J Virol* **76**: 9035–9045.
66. Kwong PD, Wyatt R, Sattentau QJ, *et al.* (2000) Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of human immunodeficiency virus. *J Virol* **74**: 1961–1972.
67. Stanfield RL, Gorny MK, Williams C, *et al.* (2004) Structural rationale for the broad neutralization of HIV-1 by human monoclonal antibody 447-52D. *Structure (Camb)* **12**: 193–204.
68. Cormier EG, Dragic T. (2002) The crown and stem of the V3 loop play distinct roles in human immunodeficiency virus type 1 envelope glycoprotein interactions with the CCR5 coreceptor. *J Virol* **76**: 8953–8957.
69. Suphaphiphat P, Thitithanyanont A, Paca-Uccaralertkun S, *et al.* (2003) Effect of amino acid substitution of the V3 and bridging sheet residues in human immunodeficiency virus type 1 subtype C gp120 on CCR5 utilization. *J Virol* **77**: 3832 –3837.
70. Cao J, Sullivan N, Desjardin E, *et al.* (1997) Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. *J Virol* **71**: 9808–9812.
71. Sharon M, Kessler N, Levy R, *et al.* (2003) Alternative conformations of HIV-1 V3 loops mimic beta hairpins in chemokines, suggesting a mechanism for coreceptor selectivity. *Structure (Camb)* **11**: 225–236.
72. Yonezawa A, Hori T, Takaori-Kondo A, *et al.* (2001) Replacement of the V3 region of gp120 with SDF-1 preserves the infectivity of T-cell

line-tropic human immunodeficiency virus type 1. *J Virol* **75**: 4258–4267.
73. Leonard CK, Spellman MW, Riddle L, *et al.* (1990) Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J Biol Chem* **265**: 10373–10382.
74. Abrignani S, Montagna D, Jeannet M, *et al.* (1990) Priming of $CD4^+$ T cells specific for conserved regions of human immunodeficiency virus glycoprotein gp120 in humans immunized with a recombinant envelope protein. *Proc Natl Acad Sci USA* **87**: 6136–6140.
75. Haigwood NL, Shuster JR, Moore GK, *et al.* (1990) Importance of hypervariable regions of HIV-1 gp120 in the generation of virus neutralizing antibodies. *AIDS Res Hum Retroviruses* **6**: 855–869.
76. Chen B, Vogan EM, Gong H, *et al.* (2005) Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature* **433**: 834–841.
77. Wyatt R, Kwong PD, Desjardins E, *et al.* (1998) The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**: 705–711.
78. Sanders RW, Vesanen M, Schuelke N, *et al.* (2002) Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1. *J Virol* **76**: 8875–8889.
79. Trkola A, Purtscher, Muster T, *et al.* (1996) Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J Virol* **70**: 1100–1108.
80. Bures R, Morris L, Williamson C, *et al.* (2002) Regional clustering of shared neutralization determinants on primary isolates of clade C human immunodeficiency virus type 1 from South Africa. *J Virol* **76**: 2233–2244.
81. Calarese DA, Scanlan CN, Zwick MB, *et al.* (2003) Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300**: 2065–2071.
82. Cao YZ, Qin LM, Zhang LQ, *et al.* (1995) Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection. *New Eng J Med* **332**: 201–208.
83. Stamatos NM, Mascola JR, Kalyanaraman VS, *et al.* (1998) Neutralizing antibodies from the sera of human immunodeficiency virus type 1-infected individuals bind to monomeric gp120 and oligomeric gp140. *J Virol* **72**: 9656–9667.

84. Fung MS, Sun CR, Gordon WL, *et al.* (1992) Identification and characterization of a neutralization site within the second variable region of human immunodeficiency virus type 1 gp120. *J Virol* **66**: 848–856.
85. Gorny MK, Moore JP, Conley AJ, *et al.* (1994) Human anti-V2 monoclonal antibody that neutralizes primary but not laboratory isolates of human immunodeficiency virus type 1. *J Virol* **68**: 8312–8320.
86. Gorny MK, Xu JY, Gianakakos V, *et al.* (1991) Production of site-selected neutralizing human monoclonal antibodies against the third variable domain of the human immunodeficiency virus type 1 envelope glycoprotein. *Proc Natl Acad Sci USA* **88**: 3238–3242.
87. Binley JM, Ditzel HJ, Barbas CF, 3rd, *et al.* (1996) Human antibody responses to HIV type 1 glycoprotein 41 cloned in phage display libraries suggest three major epitopes are recognized and give evidence for conserved antibody motifs in antigen binding. *AIDS Res Hum Retroviruses* **12**: 911–924.
88. Goudsmit J. (1988) Immunodominant B-cell epitopes of the HIV-1 envelope recognized by infected and immunized hosts. *AIDS* **2**: S41–S45.
89. Xu JY, Gorny MK, Palker T, *et al.* (1991) Epitope mapping of two immunodominant domains of gp41, the transmembrane protein of human immunodeficiency virus type 1, using ten human monoclonal antibodies. *J Virol* **65**: 4832–4838.
90. Steimer KS, Scandella CJ, Skiles PV, Haigwood NL. (1991) Neutralization of divergent HIV-1 isolates by conformation-dependent human antibodies to Gp120. *Science* **254**: 105–108.
91. Emini EA, Schleif WA, Nunberg JH, *et al.* (1992) Prevention of HIV-1 infection in chimpanzees by gp120 V3 domain-specific monoclonal antibody. *Nature* **355**: 728–730.
92. Conley AJ, Kessler JA, II, Boots LJ, *et al.* (1996) The consequence of passive administration of an anti-human immunodeficiency virus type 1 neutralizing monoclonal antibody before challenge of chimpanzees with a primary virus isolate. *J Virol* **70**: 6751–6758.
93. Mosier DE, Gulizia RJ, Baird SM, Wilson DB. (1988) Transfer of a functional human immune system to mice with severe combined immunodeficiency. *Nature* **335**: 256–259.
94. Mosier DE, Gulizia RJ, Baird SM, *et al.* (1991) Human immunodeficiency virus infection of human-PBL-SCID mice. *Science* **251**: 791–794.

95. Gauduin MC, Safrit JT, Weir R, *et al.* (1995) Pre- and postexposure protection against human immunodeficiency virus type 1 infection mediated by a monoclonal antibody. *J Infect Dis* **171**: 1203–1209.
96. Parren PW, Ditzel HJ, Gulizia RJ, *et al.* (1995) Protection against HIV-1 infection in hu-PBL-SCID mice by passive immunization with a neutralizing human monoclonal antibody against the gp120 CD4-binding site. *AIDS* **9**: F1–6.
97. Safrit JT, Fung MS, Andrews CA, *et al.* (1993) hu-PBL-SCID mice can be protected from HIV-1 infection by passive transfer of monoclonal antibody to the principal neutralizing determinant of envelope gp120. *AIDS* **7**: 15–21.
98. Parren PW, Gauduin MC, Koup RA, *et al.* (1997) Erratum to "Relevance of the antibody response against human immunodeficiency virus type 1 envelope to vaccine design". *Immunol Lett* **58**: 125–132.
99. Parren PW, Gauduin MC, Koup RA, *et al.* (1997) Erratum to "Relevance of the antibody response against human immunodeficiency virus type 1 envelope to vaccine design". *Immunol Lett* **58**: 125–132.
100. Mosier DE. (1995) Distinct rate and patterns of human $CD4^+$ T-cell depletion in hu-PBL-SCID mice infected with different isolates of the human immunodeficiency virus. *J Clin Immunol* **15**: 130S–133S.
101. Mascola JR, Lewis MG, Stiegler G, *et al.* (1999) Protection of macaques against pathogenic simian/human immunodeficiency virus 89.6PD by passive transfer of neutralizing antibodies. *J Virol* **73**: 4009–4018.
102. Mascola JR, Stiegler G, VanCott TC, *et al.* (2000) Protection of macaques against vaginal transmission of a pathogenic HIV-1/SIV chimeric virus by passive infusion of neutralizing antibodies. *Nature Medicine* **6**: 207–210.
103. Parren PW, Wang M, Trkola A, *et al.* (1998) Antibody neutralization-resistant primary isolates of human immunodeficiency virus type 1. *J Virol* **72**: 10270–10274.
104. Mascola JR, Lewis MG, Stiegler G, *et al.* (1999) Protection of macaques against pathogenic simian/human immunodeficiency virus 89.6PD by passive transfer of neutralizing antibodies. *J Virol* **73**: 4009–4018.
105. Nishimura Y, Igarashi T, Haigwood NL, *et al.* (2003) Transfer of neutralizing IgG to macaques 6 h but not 24 h after SHIV infection confers sterilizing protection: implications for HIV-1 vaccine development. *Proc Natl Acad Sci USA* **100**: 15131–15136 Epub 2003 Nov 19.

106. Hofmann-Lehmann R, Vlasak J, Rasmussen RA, *et al.* (2002) Postnatal pre- and postexposure passive immunization strategies: protection of neonatal macaques against oral simian-human immunodeficiency virus challenge. *J Med Primatol* **31**: 109–119.
107. Ruprecht RM, Hofmann-Lehmann R, Smith-Franklin BA, *et al.* (2001) Protection of neonatal macaques against experimental SHIV infection by human neutralizing monoclonal antibodies. *Transfus Clin Biol* **8**: 350–358.
108. Safrit JT, Ruprecht R, Ferrantelli F, *et al.* (2004) Immunoprophylaxis to prevent mother-to-child transmission of HIV-1. *J Acquir Immune Defic Syndr* **35**: 169–177.
109. Ferrantelli F, Hofmann-Lehmann R, Rasmussen RA, *et al.* (2003) Post-exposure prophylaxis with human monoclonal antibodies prevented SHIV89.6P infection or disease in neonatal macaques. *AIDS* **17**: 301–309.
110. Ferrantelli F, Kitabwalla M, Rasmussen RA, *et al.* (2004) Potent cross-group neutralization of primary human immunodeficiency virus isolates with monoclonal antibodies — implications for acquired immunodeficiency syndrome vaccine. *J Infect Dis* **189**: 71–74.
111. Ho DD, Sarngadharan MG, Hirsch MS, *et al.* (1987) Human immunodeficiency virus neutralizing antibodies recognize several conserved domains on the envelope glycoproteins. *J Virol* **61**: 2024–2028.
112. Montefiori DC, Pantaleo G, Fink LM, *et al.* (1996) Neutralizing and infection-enhancing antibody responses to human immunodeficiency virus type 1 in long-term nonprogressors. *J Infect Dis* **173**: 60–67.
113. Moog C, Fleury HJ, Pellegrin I, *et al.* (1997) Autologous and heterologous neutralizing antibody responses following initial seroconversion in human immunodeficiency virus type 1-infected individuals. *J Virol* **71**: 3734–3741.
114. Moore JP, Sattentau QJ, Wyatt R, Sodroski J. (1994) Probing the structure of the human immunodeficiency virus surface glycoprotein gp120 with a panel of monoclonal antibodies. *J Virol* **68**: 469–484.
115. Pilgrim AK, Pantaleo G, Cohen OJ, *et al.* (1997) Neutralizing antibody responses to human immunodeficiency virus type 1 in primary infection and long-term-nonprogressive infection. *J Infect Dis* **176**: 924–932.
116. Wrin T, Crawford L, Sawyer L, *et al.* (1994) Neutralizing antibody responses to autologous and heterologous isolates of human immunodeficiency virus. *J Acquir Immune Defic Syndr* **7**: 211–219.

117. Pilgrim AK, Pantaleo G, Cohen OJ, *et al.* (1997) Neutralizing antibody responses to human immunodeficiency virus type 1 in primary infection and long-term-nonprogressive infection. *J Infect Dis* **176**: 924–932.
118. Richman DD, Wrin T, Little SJ, Petropoulos CJ. (2003) Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc Natl Acad Sci USA* **100**: 4144–4149 Epub 2003 Mar 18.
119. Watkins BA, Buge S, Aldrich K, *et al.* (1996) Resistance of human immunodeficiency virus type 1 to neutralization by natural antisera occurs through single amino acid substitutions that cause changes in antibody binding at multiple sites. *J Virol* **70**: 8431–8437.
120. Wei X, Decker JM, Wang S, *et al.* (2003) Antibody neutralization and escape by HIV-1. *Nature* **422**: 307–312.
121. Moore JP, Ho DD. (1993) Antibodies to discontinuous or conformationally sensitive epitopes on the gp120 glycoprotein of human immunodeficiency virus type 1 are highly prevalent in sera of infected humans. *J Virol* **67**: 863–875.
122. VanCott TC, Bethke FR, Burke DS, *et al.* (1995) Lack of induction of antibodies specific for conserved, discontinuous epitopes of HIV-1 envelope glycoprotein by candidate AIDS vaccines. *J Immunol* **155**: 4100–4110.
123. Donners H, Vermoesen T, Willems B, *et al.* (2003) The first generation of candidate HIV-1 vaccines can induce antibodies able to neutralize primary isolates in assays with extended incubation phases. *Vaccine* **22**: 104–111.
124. Johnson WE, Desrosiers RC. (2002) Viral persistance: HIV's strategies of immune system evasion. *Annu Rev Med* **53**: 499–518.
125. Flynn NM, Forthal DN, Harro CD, *et al.* (2005) Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *J Infect Dis* **191**: 654–665.
126. Francis DP, Heyward WL, Popovic V, *et al.* (2003) Candidate HIV/AIDS vaccines: lessons learned from the World's first phase III efficacy trials. *AIDS* **17**: 147–156.
127. Gilbert PB, Peterson ML, Follmann D, *et al.* (2005) Correlation between immunologic responses to a recombinant glycoprotein 120 vaccine and incidence of HIV-1 infection in a phase 3 HIV-1 preventive vaccine trial. *J Infect Dis* **191**: 666–677.
128. Fouts TR, Binley JM, Trkola A, *et al.* (1997) Interaction of polyclonal and monoclonal anti-glycoprotein 120 antibodies with oligomeric

glycoprotein 120-glycoprotein 41 complexes of a primary HIV type 1 isolate relationship to neutralization. AIDS Research and *Human Retroviruses* **14**: 591–597.

129. Sattentau QJ, Moore JP. (1995) Human immunodeficiency virus type 1 neutralization is determined by epitope exposure on the gp120 oligomer. *J Exp Med* **182**: 185–196.
130. VanCott TC, Mascola JR, Kaminski RW, *et al.* (1997) Antibodies with specificity to native gp120 and neutralization activity against primary human immunodeficiency virus type 1 isolates elicited by immunization with oligomeric gp160. *J Virol* **71**: 4319–4330.
131. Binley JM, Sanders RW, Clas B, *et al.* (2000) A recombinant human immunodeficiency virus type 1 envelope glycoprotein complex stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits is an antigenic mimic of the trimeric virion-associated structure. *J Virol* **74**: 627–643.
132. Broder CC, Earl PL, Long D, *et al.* (1994) Antigenic implications of human immunodeficiency virus type 1 envelope quaternary structure: oligomer-specific and -sensitive monoclonal antibodies. *Proc Natl Acad Sci USA* **91**: 11699–11703.
133. Earl PL, Broder CC, Long D, *et al.* (1994) Native oligomeric human immunodeficiency virus type 1 envelope glycoprotein elicits diverse monoclonal antibody reactivities. *J Virol* **68**: 3015–3026.
134. Earl PL, Koenig S, Moss B. (1991) Biological and immunological properties of human immunodeficiency virus type 1 envelope glycoprotein: analysis of proteins with truncations and deletions expressed by recombinant vaccinia viruses. *J Virol* **65**: 31–41.
135. Earl PL, Sugiura W, Montefiori DC, *et al.* (2001) Immunogenicity and protective efficacy of oligomeric human immunodeficiency virus type 1 gp140. *J Virol* **75**: 645–653.
136. Farzan M, Choe H, Desjardins E, *et al.* (1998) Stabilization of human immunodeficiency virus type 1 envelope glycoprotein trimers by disulfide bonds introduced into the gp41 glycoprotein ectodomain. *J Virol* **72**: 7620–7625.
137. Sanders RW, Schiffner L, Master A, *et al.* (2000) Variable-loop-deleted variants of the human immunodeficiency virus type 1 envelope glycoprotein can be stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits. *J Virol* **74**: 5091–5100.
138. Schulke N, Vesanen MS, Sanders RW, *et al.* (2002) Oligomeric and conformational properties of a proteolytically mature, disulfide-stabilized

human immunodeficiency virus type 1 gp140 envelope glycoprotein. *J Virol* **76**: 7760–7776.

139. Srivastava IK, Stamatatos L, Kan E, *et al.* (2003) Purification, characterization, and immunogenicity of a soluble trimeric envelope protein containing a partial deletion of the V2 loop derived from SF162, an R5-tropic human immunodeficiency virus type 1 isolate. *J Virol* **77**: 11244–11259.
140. Srivastava IK, Stamatatos L, Legg H, *et al.* (2002) Purification and characterization of oligomeric envelope glycoprotein from a primary r5 subtype B human immunodeficiency virus. *J Virol* **76**: 2835–2847.
141. Yang X, Farzan M, Wyatt R, Sodroski J. (2000) Characterization of stable, soluble trimers containing complete ectodomains of human immunodeficiency virus type 1 envelope glycoproteins. *J Virol* **74**: 5716–5725.
142. Yang X, Lee J, Mahony EM, *et al.* (2002) Highly stable trimers formed by human immunodeficiency virus type 1 envelope glycoproteins fused with the trimeric motif of T4 bacteriophage fibritin. *J Virol* **76**: 4634–4642.
143. Yang X, Wyatt R, Sodroski J. (2001) Improved elicitation of neutralizing antibodies against primary human immunodeficiency viruses by soluble stabilized envelope glycoprotein trimers. *J Virol* **75**: 1165–1171.
144. Barnett SW, Lu S, Srivastava I, *et al.* (2001) The ability of an oligomeric human immunodeficiency virus type 1 (HIV-1) envelope antigen to elicit neutralizing antibodies against primary HIV-1 isolates is improved following partial deletion of the second hypervariable region. *J Virol* **75**: 5526–5540.
145. Srivastava IK, VanDorsten K, Vojtech L, *et al.* (2003) Changes in the immunogenic properties of soluble gp140 human immunodeficiency virus envelope constructs upon partial deletion of the second hypervariable region. *J Virol* **77**: 2310–2320.
146. Cherpelis S, Jin X, Gettie A, *et al.* (2001) DNA-immunization with a V2 deleted HIV-1 envelope elicits protective antibodies in macaques. *Immunol Lett* **79**: 47–55.
147. Cherpelis S, Shrivastava I, Gettie A, *et al.* (2001) DNA vaccination with the human immunodeficiency virus type 1 SF162DeltaV2 envelope elicits immune responses that offer partial protection from simian/human immunodeficiency virus infection to CD8(+) T-cell-depleted rhesus macaques. *J Virol* **75**: 1547–1550.

148. Buckner C, Gines LG, Saunders CJ, *et al.* (2004) Priming B cell-mediated anti-HIV envelope responses by vaccination allows for the long-term control of infection in macaques exposed to a R5-tropic SHIV. *Virology* **320**: 167–180.
149. Quinnan GV, Jr., Yu XF, Lewis MG, *et al.* (2005) Protection of rhesus monkeys against infection with minimally pathogenic simian-human immunodeficiency virus: correlations with neutralizing antibodies and cytotoxic T cells. *J Virol* **79**: 3358–3369.
150. Lian Y, Srivastava I, Gomez-Roman VR, *et al.* (2005) Evaluation of envelope vaccines derived from the South African subtype C human immunodeficiency virus type 1 TV1 strain. *J Virol* **79**: 13338–13349.
151. Myszka DG, Sweet RW, Hensley P, *et al.* (2000) Energetics of the HIV gp120-CD4 binding reaction. *Proc Natl Acad Sci USA* **97**: 9026–9031.
152. Pal AK, Rahman S, Chatterjee SN. (1993) Electron microscopic study of phages and aberrant structures produced by induction of prophage kappa in Vibrio cholerae el tor cells. *Indian J Exp Biol* **31**: 955–962.
153. Gershoni J, Denisova G, Raviv D, *et al.* (1993) HIV binding to its receptor creates specific epitopes for CD4/gp120 complex. *FEBS J* **7**: 1185–1187.
154. Xiang SH, Wang L, Abreu M, *et al.* (2003) Epitope mapping and characterization of a novel CD4-induced human monoclonal antibody capable of neutralizing primary HIV-1 strains. *Virology* **315**: 124–134.
155. DeVico AL, Rahman R, Welch J, *et al.* (1995) Monoclonal antibodies raised against covalently crosslinked complexes of human immunodeficiency virus type 1 gp120 and CD4 receptor identify a novel complex-dependent epitope on gp 120. *Virology* **211**: 583–588.
156. Srivastava I, Ulmer J, Barnett S. (2004) Neutralizing antibody responses to HIV: role in protective immunity and challenges for vaccine design. *Expert Opinion in Vaccines.*
157. Kwong PD, Wyatt R, Robinson J, *et al.* (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**: 648–659.
158. Clackson T, Wells JA. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science* **267**: 383–386.
159. Vita C. (1997) Engineering novel proteins by transfer of active sites to natural scaffolds. *Curr Opin Biotechnol* **8**: 429–434.

160. Vita C, Vizzavona J, Drakopoulou E, *et al.* (1998) Novel miniproteins engineered by the transfer of active sites to small natural scaffolds. *Biopolymers* **47**: 93–100.
161. Vita C, Drakopoulou E, Vizzavona J, *et al.* (1999) Rational engineering of a miniprotein that reproduces the core of the CD4 site interacting with HIV-1 envelope glycoprotein. *Proc Natl Acad Sci USA* **96**: 13091–13096.
162. Martin L, Stricher F, Misse D, *et al.* (2003) Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nat Biotechnol* **21**: 71–76.
163. Fouts T, Godfrey K, Bobb K, *et al.* (2002) Crosslinked HIV-1 envelope-CD4 receptor complexes elicit broadly cross-reactive neutralizing antibodies in rhesus macaques. *Proc Natl Acad Sci USA* **99**: 11842–11847.
164. Fouts TB, Godfrey K, Syed A, *et al.* (2003) Structural Characterization and Immunogenicity of a Single Chain Polypeptide Analogue of the HIV-1 gp120-CD4 Receptor Complex Using a CD4 Mimic Protein, CD4M9. *AIDS Vaccine* **56** (Abstract 39).
165. Guo Q, Ho HT, Dicker I, *et al.* (2003) Biochemical and genetic characterizations of a novel human immunodeficiency virus type 1 inhibitor that blocks gp120-CD4 interactions. *J Virol* **77**: 10528–10536.
166. Kwong PD, Wyatt R, Robinson J, *et al.* (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**: 648–659.
167. Huang CC, Tang M, Zhang MY, *et al.* (2005) Structure of a V3-containing HIV-1 gp120 core. *Science* **310**: 1025–1028.
168. Malenbaum SE, Yang D, Cavacini L, *et al.* (2000) The N-terminal V3 loop glycan modulates the interaction of clade A and B human immunodeficiency virus type 1 envelopes with CD4 and chemokine receptors. *J Virol* **74**: 11008–11016.
169. Reitter JN, Means RE, Desrosiers RC. (1998) A role for carbohydrates in immune evasion in AIDS. *Nat Med* **4**: 679–684.
170. McCaffrey RA, Saunders C, Hensel M, Stamatatos L. (2004) N-linked glycosylation of the V3 loop and the immunologically silent face of gp120 protects human immunodeficiency virus type 1 SF162 from neutralization by anti-gp120 and anti-gp41 antibodies. *J Virol* **78**: 3279–3295.
171. Pantophlet R, Ollmann Saphire E, Poignard P, *et al.* (2003) Fine mapping of the interaction of neutralizing and nonneutralizing monoclonal

antibodies with the CD4 binding site of human immunodeficiency virus type 1 gp120. *J Virol* **77**: 642–658.
172. Pantophlet R, Wilson IA, Burton DR. (2003) Hyperglycosylated mutants of human immunodeficiency virus (HIV) type 1 monomeric gp120 as novel antigens for HIV vaccine design. *J Virol* **77**: 5889–5901.
173. Stamatatos L, Wiskerchen M, Cheng-Mayer C. (1998) Effect of major deletions in the V1 and V2 loops of a macrophage-tropic HIV type 1 isolate on viral envelope structure, cell entry, and replication. *AIDS Res Hum Retroviruses* **14**: 1129–1139.
174. Stamatatos L, Lim M, Cheng-Mayer C. (2000) Generation and structural analysis of soluble oligomeric gp140 envelope proteins derived from neutralization-resistant and neutralization-susceptible primary HIV type 1 isolates [In Process Citation]. *AIDS Res Hum Retroviruses* **16**: 981–994.
175. Zwick MB, Komori HK, Stanfield RL, *et al.* (2004) The long third complementarity-determining region of the heavy chain is important in the activity of the broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2F5. *J Virol* **78**: 3155–3161.
176. LaCasse RA, Follis KE, Trahey M, *et al.* (1999) Fusion-competent vaccines: broad neutralization of primary isolates of HIV. *Science* **283**: 357–362.
177. Wild C, Oas T, McDanal C, *et al.* (1992) A synthetic peptide inhibitor of human immunodeficiency virus replication: correlation between solution structure and viral inhibition. *Proc Natl Acad Sci USA* **89**: 10537–10541.
178. Zanetti G, Briggs JA, Grunewald K, *et al.* (2006) Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. *PLoS Pathog* **2**: e83.
179. Zhu P, Liu J, Bess J, Jr., *et al.* (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441**: 847–852.
180. Mascola JR, Montefiori DC. (2003) HIV-1: nature's master of disguise. *Nat Med* **9**: 393–394.
181. Cole KS, Alvarez M, Elliott DH, *et al.* (2001) Characterization of neutralization epitopes of simian immunodeficiency virus (SIV) recognized by rhesus monoclonal antibodies derived from monkeys infected with an attenuated SIV strain. *Virology* **290**: 59–73.
182. Sattentau QJ, Moore JP. (1991) Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J Exp Med* **174**: 407–415.

183. Sattentau QJ, Moore JP, Vignaux F, *et al.* (1993) Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. *J Virol* **67**: 7383–7393.
184. Salzwedel K, West JT, Hunter E. (1999) A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity. *J Virol* **73**: 2469–2480.
185. Caffrey M, Cai M, Kaufman J, *et al.* (1998) Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. *EMBO J* **17**: 4572–4584.
186. Lu S, Wang S, Mascola J, *et al.* (2004) Presented at the Keystone Symposium on HIV Vaccine Development: Progress and Prospects; Abstract # 245. Whistler, British Columbia, Canada.
187. Sambor A, Louder M, Welcher B, *et al.* (2004) Presented at the Keystone Symposium on HIV Vaccine Development: Progress and Prospects; Abstract # 306. Whistler, British Columbia, Canada.
188. Mascola JR, Stiegler G, VanCott TC, *et al.* (2000) Protection of macaques against vaginal transmission of a pathogenic HIV-1/SIV chimeric virus by passive infusion of neutralizing antibodies. *Nat Med* **6**: 207–210.
189. Parren PW, Marx PA, Hessell AJ, *et al.* (2001) Antibody protects macaques against vaginal challenge with a pathogenic R5 simian/human immunodeficiency virus at serum levels giving complete neutralization *in vitro*. *J Virol* **75**: 8340–8347.

This page intentionally left blank

Chapter 14

Insights into the *Caliciviridae* Family

*Grant Hansman**

Introduction

The family *Caliciviridae* contains four genera, *Norovirus*, *Sapovirus*, *Lagovirus*, and *Vesivirus*, which include norovirus (NoV), sapovirus (SaV), rabbit hemorrhagic disease virus (RHDV), and feline calicivirus (FCV) strains, respectively. Human NoV, previously known as small-round structured viruses (SRSV) or Norwalk-like viruses (NLVs), is the leading cause of gastroenteritis in the world. The prototype strain of human NoV, the Norwalk virus (Hu/NV/Norwalk virus/1968/US), was first discovered from an outbreak of gastroenteritis in an elementary school in Norwalk, Ohio, USA, in 1968.[1] Human SaV strains are thought to mostly infect infants, occasionally causing outbreaks of gastroenteritis.[2–4] The prototype strain of human SaV, the Sapporo virus (Hu/SV/Sapporo virus/1977/JP), was originally discovered from an outbreak in a home for infants in Sapporo, Japan, in 1977.[5] Most animal caliciviruses are grouped within the other two genera. *Vesivirus* and *Lagovirus* cause a variety of diseases, such as gastroenteritis, vesicular lesions, respiratory infections, reproductive failure, and hemorrhagic disease.

*National Institute of Infectious Diseases, Tokyo.

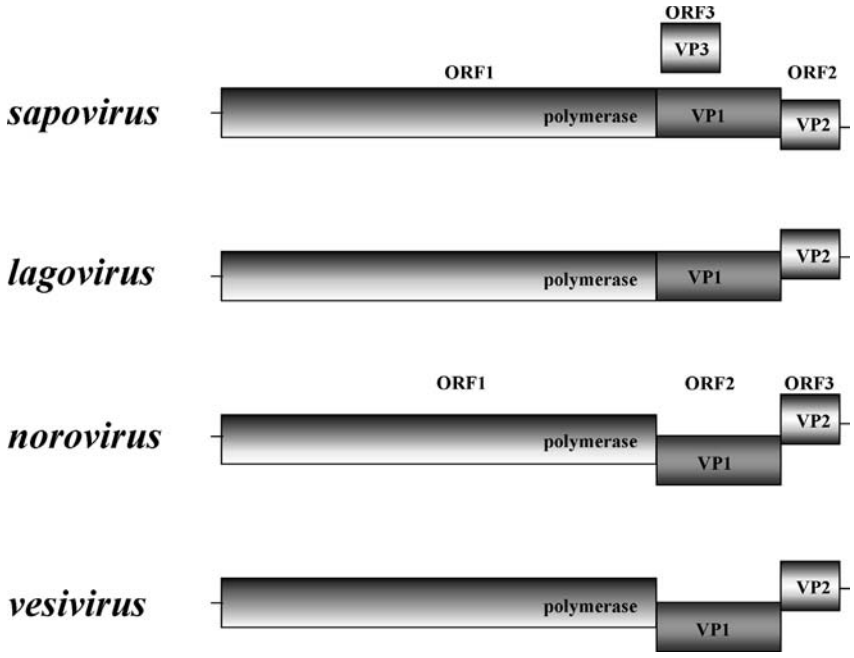


Fig. 1. Schematics of the family *Caliciviridae*. For *Norovirus* and *vesivirus*, ORF1 encodes the non-structural proteins, ORF2 encodes the major capsid protein (VP1), and ORF3 encodes a small protein (VP2). For *sapovirus* and *lagovirus*, ORF1 encodes the non-structural proteins and VP1, whereas ORF2 encodes proteins of yet-unknown functions, and ORF3 (an unknown protein) is only predicted for SaV GI, GIV, and GV genomes.

Genome

The calicivirus genome consists of a positive-sense, single-stranded RNA of approximately 7.4 to 8.3 kb in length that is organized in either two or three open reading frames (ORFs) (Fig. 1). NoV and FCV have a similar genomic organization in that ORF1 encodes non-structural proteins, including N-terminal protein, NTPase, 3A-like protein, genome-linked viral protein (VPg), 3C-like protease, and RNA-dependent RNA polymerase (RdRp). The NoV and FCV ORF2 encodes the major capsid protein (VP1) and the NoV and FCV ORF3 encodes a small protein (VP2). The NoV VP2 is associated

with VP1 stability and FCV VP2 is essential for the production of infectious virions.[6,7] The SaV GI, GIV, and GV genomes are predicted to each contain three main ORFs, whereas the SaV GII and GIII genomes (and RHDV genomes) each have only two main ORFs.[8] The SaV and RHDV ORF1 encode non-structural proteins and the capsid protein. The SaV and RHDV ORF2 and SaV ORF3 encode proteins of yet-unknown functions.

Genetic Analysis and Recombination

Human NoV can be divided into two genetically distinct genogroups (GI and GII), which can be subdivided into at least 14 GI and 17 GII clusters or genotypes.[9] Based on the capsid gene sequence, SaV can be grouped into five distinct genogroups (GI to GV).[8] Human SaV belong to GI, GII, GIV, and GV, whereas pig SaV belongs to GIII. In both NoV and SaV, the genotypes are generally maintained across the ORFs. However, a number of NoV and SaV strains failed to maintain their sequence identities for RdRp and VP1, and they were shown to be recombinant.[10–16] Evidence suggested that the recombination site occurred at the polymerase and capsid junction on ORF1. Phylogenetic analysis of the non-structural region (i.e. genome start to capsid start) grouped these strains into one cluster, while the structural region (i.e. capsid start to genome end) grouped these strains into two distinct clusters. Figure 2 shows a possible mechanism for NoV and SaV recombination.[10] Nucleotide sequence at the polymerase and capsid junction generally is conserved among human NoV and SaV strains,[14,17] likely facilitating a recombination event when nucleic acid sequences of parental strains come into physical contact in infected cells, e.g. during copy choice recombination. These data suggested that these viruses could evade host immunity by readily changing their structural region, i.e. the capsid protein that is immuno-reactive.

Expression in Insect Cells and Virus-like Particles

To date, all laboratory efforts to cultivate human NoV have failed,[18] but expression of the recombinant VP1 (rVP1) in insect or mammalian

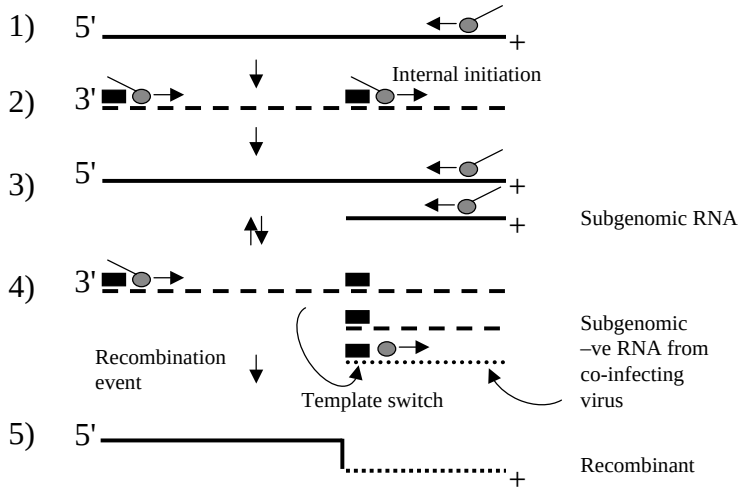
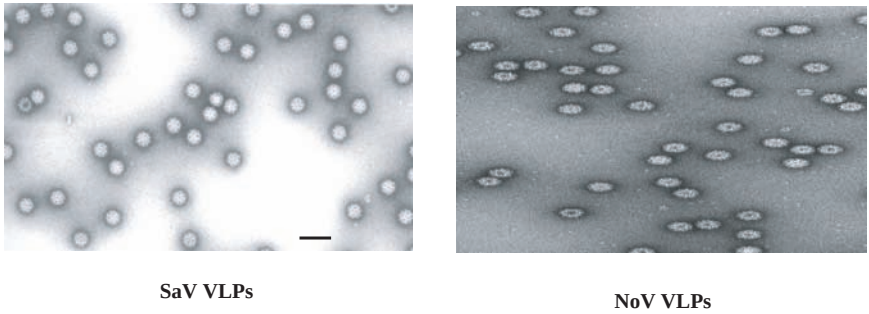


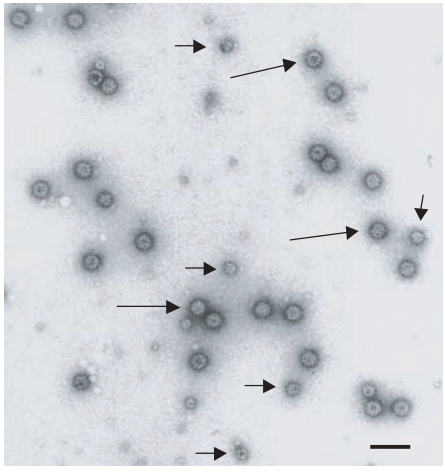
Fig. 2. A simple mechanism for recombination in NoV. 1) RNA transcription by the RNA-dependent RNA polymerase (RdRp) (grey circle) generates a negative strand intermediate (dashed line). 2) Binding of the RdRp to the almost identical RNA promoter sequences (filled boxes) generates positive-stranded (straight line) genomes and subgenomic RNA. 3) These templates direct RNA synthesis from the 3′ end that leads to the generation of both a full-length negative genome and a negative subgenomic RNA species. 4) Recombination occurs when the enzyme initiates positive strand synthesis at the 3′ end of the full-length negative strand, stalls at the subgenomic promoter, and then template switches to an available negative subgenomic RNA species generated by a co-infecting virus. The net result is a recombinant virus that has acquired new ORF2 and ORF3 sequences.

cells can result in the formation of virus-like particles (VLPs) that are morphologically similar to native NoV[6,19–24] (Fig. 3A). These studies have provided valuable information, including the high-resolution atomic structure, antigenic analysis, and binding factors. However, it remains unclear how similar these VLPs are to native virions.

Barcena *et al.*[25] showed that RHDV N-terminal-deleted rVP1 constructs could form small and native-size VLPs and that all RHDV C-terminal-deleted rVP1 constructs failed to form VLPs. In contrast, Bertolotti-Ciarlet *et al.*[26] showed that both NoV N- and



(A)



(B)

Fig. 3. (**A**) EM images of CsCl purified SaV and NoV native-size VLPs negative-stained with 2% uranyl acetate (pH 4). (**B**) EM image of CsCl purified SaV VLPs showing both small and native-size VLPs negative-stained with 4% uranyl acetate (pH 4). The long arrows show the native-size VLPs, and short arrows show the small VLPs. The bar indicates 100 nm.

C-terminal-deleted rVP1 constructs could form VLPs. SaV expression was similar to those reported in the RHDV N- and C-terminal-deleted rVP1 expression study, but were distinct from those reported in the NoV N- and C-terminal-deleted rVP1 expression study.[32] For SaV, only proteins derived from N-terminal-deleted rVP1 constructs

that began 49 nucleotides downstream assembled into VLPs, which included both small and native-size VLPs (Fig. 3B). Small virions have not yet been described in clinical stool specimens, which may suggest that these smaller VLPs were exclusive to insect cell expression.

Capsid Structure

All caliciviruses process a single capsid protein (VP1) with apparent molecular weights ranging from 59 to 65 kDa. The VP1 is responsible for assembly, host interactions, and immunogenicity. Capsids that are made of a single-structure protein are unusual among animal viruses, but are common among plant viruses. The cryo-EM structure of the Norwalk virus VLPs was determined in 1994.[27] The Norwalk virus capsid was found to exhibit a T = 3 icosahedral symmetry with 180 molecules of the VP1 organized into 90 dimeric capsomers. The X-ray crystallographic structure of Norwalk virus capsid was determined in 1999.[28] The Norwalk virus capsid was found to have two principal domains: a shell (S) domain and a protrusion (P) domain, the latter of which is further divided into two subdomains called P1 and P2 (Fig. 4). The S domain is well conserved among the caliciviruses, whereas the P1 domain is moderately conserved and the P2 domain is highly variable. The P domain is thought to be exclusively involved in dimeric interactions for all caliciviruses; furthermore, the modular S-P1-P2 domain organization of the capsid protein and the tertiary structure of the P1 subdomain appear to be conserved in all caliciviruses.[29] Particle assembly is not linked to protein-RNA interactions because VLP formation occurs without the genomic RNA. The 3D structure of vesivirus virions (containing RNA), empty lagovirus VLPs, and empty SaV VLPs have been determined by cryo-EM to a resolution of ~22 Å.[29–31] In addition, NoV GII VLPs were determined. Comparative studies found notable surface differences, though similar general features were also observed. The P2 subdomain had the most variation. In all caliciviruses, the shell is formed by the S domain, which is located between radii 100 and 150 Å. However, the vesivirus contains another an inner contiguous shell (IS) domain, though this may be associated with RNA. For NoV, the P2 subdomain was described as a rectangular platform; for SaV, the P2 subdomain was

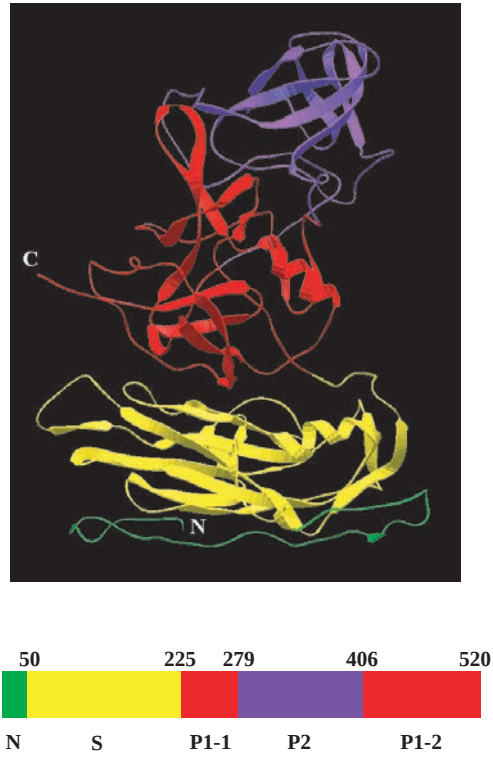


Fig. 4. The VP1 domains and the ribbon structure of Norwalk virus (accession number M87661) developed using the Swiss model software (www.expasy.org/swissmod/SWISS-MODEL.html).

described as saddle-shaped; and for SMSV4, the P2 subdomain was described as parallelogram-shaped with a dent at the center.[29]

VP1: N Terminus

The N-terminus comprises of the first 50 residues of the capsid protein (Fig. 4). This is located in the interior of the particles. The N terminus of NoV VP1 is acidic and is not thought to interact directly with the viral genome.[6] Several studies have found that deletion of the small part of the N terminus of NoV, SaV, or RHDV did not affect

VLP formation.[25,26,32] However, further deletion of the N terminus affected assembly, indicating that the N terminal region is a molecular switch that contains determinates for assembly.

VP1: S Domain

The contiguous shell is formed from the closely interacting S domains of the 180 capsid proteins. The S domain is composed of residues 50 to 225 and is relatively conserved among caliciviruses. In the Norwalk virus, the S domain folds into eight-strand antiparallel β-barrel structures.[28] This eight-strand antiparallel β-barrel is a common central structural motif in the T = 3 plant and insect viruses. One study suggested that all caliciviruses exhibit this eight-strand antiparallel β-barrel structure and that it is the most conserved structural feature in the calicivirus structures.[29] And another study showed that expression of the S domain alone resulted in the formation of smooth particles, but deletion of part of the S domain resulted in the formation of unstable particles, indicating that the S domain contained all the determinates, or a scaffold, for assembly.[6] The Norwalk virus shell lies between radii of 110 and 150 Å.

VP1: P1 Subdomain

The S domain is connected to the P1-1 by a flexible hinge. For the Norwalk virus, the P1 subdomain consists of residues 226 to 278 (called P1-1) and 406 to 520 (called P1-2). The P1-1 subdomain contains three β strands, whereas the P1-2 subdomain contains six β strands and an α helix.[28] Comparison studies of NoV genogroup I and II VLP cryo-EM structures found that the P1-1 and P2 subdomains were involved in the intradimeric interactions, whereas for SaV and vesivirus, only the P2 subdomain was thought to be involved in the intradimeric interactions and the P1 domain was involved in the interdimeric interaction.[29] Amino acid alignments, secondary structure predictions, and fitting models have indicated that the P1 subdomain is conserved in all caliciviruses.[29]

VP1: P2 Subdomain

The P2 subdomain has been described as a large insertion, distal globular portion of the arch, and a replaceable module. Amino acid alignments have shown that the P2 subdomain is the most variable region. In addition, a large number of insertions were seen in two genera, *Sapovirus* and *Vesiviruis*, with respect to *Norovirus*. The P2 subdomain occupies the residues 279 to 405 of NoV VP1. The P2 residues 285 to 380 fold into a compact barrel-like structure consisting of six β strands.[28] The fold of this subdomain was found to be similar to the elongation factor-Tu (EF-Tu). The EF-Tu is a GTP binding protein involved in transporting aminoactyl-tRNAs to ribosomes, which suggests a possible role in viral or cellular RNA translation and regulation of protein synthesis. In addition, the P2 subdomain is thought to contain the determinants of strain specificity, cell binding, and antigencity. For example, monoclonal antibodies that recognize regions in the P2 subdomain inhibit binding of NoV VLPs to cells.[33,34]

Subunits

In order for NoV to form a T = 3 icosahedral structure, the capsid protein forms three quasiequivalent positions, termed A, B, and C.[28] The structures of the A and B subunits are similar, maintaining relative orientations between the S and P domains. The S domain is involved in the icosahedral contacts, whereas the P domain is involved in the dimeric contacts. The dimers on the local twofolds axes are formed by A and B subunits, whereas those on the strict axes are formed by two C subunits. These three subunits form dimers, denoted AB and CC. The S domains of the AB dimer have a "bent" confirmation, whereas the S domains of the CC dimer have a "flat" confirmation. These confirmations allow for the formation of a closed shell. The AB dimers surround the icosahedral fivefolds axes, whereas the AB and CC dimers alternate around the icosahedral threefold axes.

VP2 and VP3

The highly basic VP2 may serve a function similar to the basic N terminus of plant virus capsid proteins, which are involved in RNA binding, since the N terminus of NoV VP1 is acidic and is not thought to interact directly with the viral genome.[6] However, VP2 is not essential for the expression of VLPs,[35] but may serve a regulatory function. This is thought to confer stability to the capsid protein.[6] This protein is believed to be associated with encapsulation of the genomic RNA. The function(s) of the VP3 is unknown and is only found in SaV GI, GIV, and GV strains.

Antigenicity

Expression of VP1 results in the formation of VLPs that are morphologically and antigenically similar to native viruses.[11] At least two kinds of assays have been used to examine the cross-reactivities among these VLPs, an antibody enzyme-linked immunosorbent assay (ELISA) and an antigen ELISA.[9,36–39] The antibody ELISA is broadly reactive, whereas the antigen ELISA is specific, only detecting strains that are closely related (>95% identity in the RNA polymerase region). For example, in an antibody ELISA, NoV GI/11 antiserum showed broad-range cross-reactivities, detecting 2 NoV GI and 10 GII genotypes.[35] Likewise, NoV GII/1, GII/10, and GII/12 antisera also showed broad-range cross-reactivities in a antibody ELISA, detecting several other distinct NoV GII genotypes.[35] Amino acid alignment of NoV VP1 suggested that these broad-range cross-reactivities were due to conserved amino acid residues located within the shell and/or P1-1 domains. On the other hand, unusual cross-reactivities among different GII/3 antisera were observed in an antibody ELISA. The four kinds of GII/3 antisera (strains 809, Sh5, 18-3, and 336) cross-reacted moderately to weakly against GII/3 1152 VLPs (i.e. up to eight-fold lower than homologous VLP titer). Amino acid alignments of these five GII/3 sequences revealed 1152 had three unique amino acid residues compared to the other four GII/3 sequences (amino acid positions Thr-285, Ile-372, and Ser-508), two of which were located within the P2 domain (Thr-285 and Ile-372). Amino acid secondary structure

predictions made using the PSIPRED secondary structural prediction software revealed that the VP1 secondary structures for 809, Sh5, 18-3, and 336 had a helix structure at the first unique residue, but this helix structure was absent for 1152 (Fig. 5). This helix structure may have in

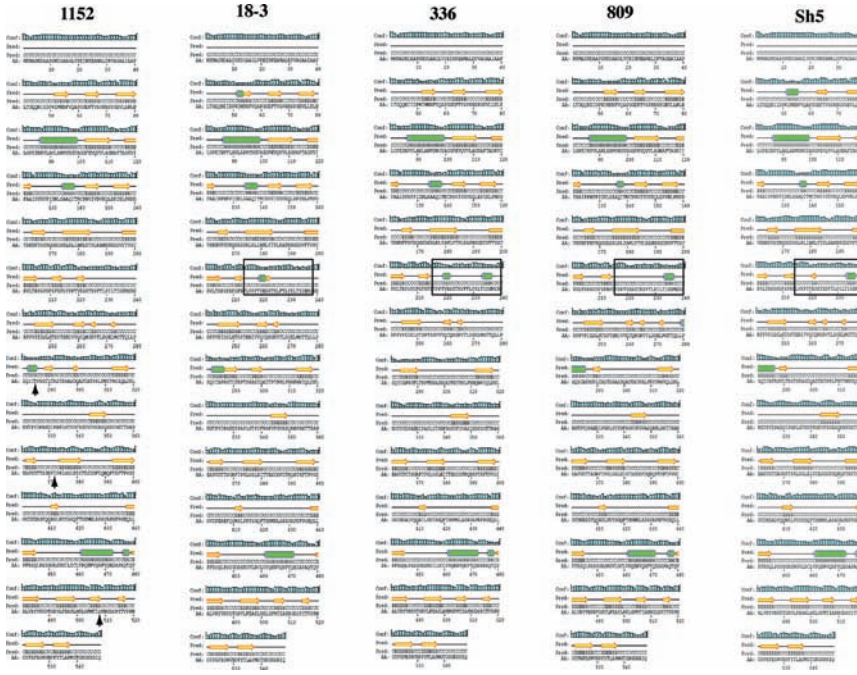


Fig. 5. Schematic representations of the complete predicted secondary structure of NoV (GII/3) 1152, 18-3, 336, 809, and Sh5 VP1. The level of confidence of prediction (Conf:) is shown on the first line, where a tall box represents a high confidence of prediction and a low box represents a low confidence of prediction. The predicted secondary structure (Pred:) is shown on the second line, where helix = cylinder, β-strand = arrow, and a line = coil. The third line also shows the predicted secondary structure (Pred:), where H = helix, E = β-strand, and C = coil. The amino acid sequence is shown on the last line (AA:). The box regions on 18-3, 336, 809, and Sh5 VP1 indicate a helix structure, which is absent on 1152 VP1. The unique amino acid residues for 1152 sequence when compared to the other four GII/3 sequences are indicated with an arrow.

part influenced the cross-reactivity among the GII/3 VLPs (i.e. without the helix structure, GII/3 1152 VLPs cross-react weakly against the other four GII/3 antisera). This suggestion may also explain NoV virulence in which some strains appear to infect a certain population over an extended period of time.[40,41] In a recent report, single amino acid changes were suggested to represent a possible way for the virus to evade the host immunity.[40] In addition, one report suggested a change in the VP1 secondary structure (i.e. a disappearance of a helix structure) was responsible for a chronic NoV infection in an immunocompromised patient for over two years.[42] Furthermore, the helix structure reported by Nilsson *et al.* was located near the helix structure described in Fig. 5, signifying an important antigenic site.

Proteolytic and Replication

Proteolytic and replication studies on human caliciviruses have been enigmatic due to their inability to grow in conventional cell cultures. Development of a self-replicating complete VLP (i.e. containing RNA) would greatly enhance the understanding of infectivity, antigenicity, and binding factors. Nevertheless, a number of important proteolytic and replication studies have been conducted. Proteolytic processing of the ORF1 polyprotein is a common feature of the caliciviruses.[43] The cleavage sites have been mapped in detail in RHDV,[44–48] FCV,[49–51] and NoV,[52–59] and more recently the cleavage map of a SaV was determined.[60] Site-directed mutagenesis demonstrated that the P1 amino acid (i.e. the amino acid immediately upstream of the scissile bond) plays a critical role in the proteolytic processing. The importance of phenylalanine at the P4 position to achieve efficient cleavage between N-terminal protein and NTPase of NoV was noted by Hardy *et al.*[54] Similarly, a phenylalanine residue was found at the P4 position in three of five cleavage sites in the ORF1 and at one cleavage site in the ORF2 of FCV. Although the details of the calicivirus genome replication, transcription, and translation remain unclear, the translation might require a cap or a cap-like structure, or VPg attached to the 5′ end of the genome and subgenome as reported in FCV.[61,62] Recently,

the interaction of NoV VPg with the translation initiation factor eIF3 has been reported,[63] and thus it is possible that the NoV VPg regulates transcription and translation initiation events. For some animal caliciviruses, synthesis of the capsid protein is initiated from the subgenomic RNA and involves sequences conserved at the 5′ end of the genomic and subgenomic RNA,[61,64] which may also be the case for NoV.[65] A number of studies have expressed human NoV in cultured human embryonic kidney 293T cells using a number of different constructs designed to produce RNA identical to the genomic RNA.[66,67] In these studies the negative-strand genomic-like RNA and the subgenomic-like RNA were transcribed by the functional viral RdRp. This was evident from the fact that a disrupted RdRp failed to produce genomic and subgenomic RNA. However, VLPs were not formed when the full-length genomic-like construct was transfected alone, but were formed when both the full-length genomic-like and subgenomic-like construct were co-transfected. NoV RNA was not packed into the VLPs when the subgenomic-like construct was expressed alone, but NoV genomic RNA appeared to be packed into the VLPs when the full-length genomic-like and subgenomic-like constructs were co-transfected. This suggested that RNA packaging signals resided in the ORF1 region and/or non-structural proteins were required for the RNA packaging and viral replication.[7,66,68,69] Recently, a murine NoV cell culture system has been established[70] of which two aspects were examined, the expression of a subgenomic RNA and the rearrangement of intracellular membranes during replication. Subgenomic RNA was detected from the murine NoV-infected cells and reorganization of intracellular membranes and loss of an intact Golgi apparatus was observed.[70] Comparisons between human NoV replication systems[66,67] and the murine NoV cell culture system[70] should provide some fundamental mechanisms.

References

1. Kapikian AZ, Wyatt RG, Dolin R, *et al.* (1972) Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. *J Virol* **10**: 1075–1081.

2. Johansson PJ, Bergentoft K, Larsson PA, *et al.* (2005) A nosocomial sapovirus-associated outbreak of gastroenteritis in adults. *Scand J Infect Dis* **37**: 200–204.
3. Noel JS, Liu BL, Humphrey CD, *et al.* (1997) Parkville virus: a novel genetic variant of human calicivirus in the Sapporo virus clade, associated with an outbreak of gastroenteritis in adults. *J Med Virol* **52**: 173–178.
4. Okada M, Shinozaki K, Ogawa T, Kaiho I. (2002) Molecular epidemiology and phylogenetic analysis of Sapporo-like viruses. *Arch Virol* **147**: 1445–1451.
5. Chiba S, Sakuma Y, Kogasaka R, *et al.* (1979) An outbreak of gastroenteritis associated with calicivirus in an infant home. *J Med Virol* **4**: 249–254.
6. Bertolotti-Ciarlet A, Crawford SE, Hutson AM, Estes MK. (2003) The 3′ End of Norwalk Virus mRNA Contains Determinants That Regulate the Expression and Stability of the Viral Capsid Protein VP1: a Novel Function for the VP2 Protein. *J Virol* 77: 11603–11615.
7. Sosnovtsev SV, Belliot G, Chang KO, *et al.* (2005) Feline calicivirus VP2 is essential for the production of infectious virions. *J Virol* **79**: 4012–4024.
8. Farkas T, Zhong WM, Jing Y, *et al.* (2004) Genetic diversity among sapoviruses. *Arch Virol* **149**: 1309–1323.
9. Kageyama T, Shinohara M, Uchida K, *et al.* (2004) Coexistence of multiple genotypes, including newly identified genotypes, in outbreaks of gastroenteritis due to Norovirus in Japan. *J Clin Microbiol* **42**: 2988–2995.
10. Bull RA, Hansman GS, Clancy LE, *et al.* (2005) Norovirus recombination in ORF1/ORF2 overlap. *Emerg Infect Dis* **11**: 1079–1085.
11. Hansman GS, Doan LT, Kguyen TA, *et al.* (2004) Detection of norovirus and sapovirus infection among children with gastroenteritis in Ho Chi Minh City, Vietnam. *Arch Virol* **149**: 1673–1688.
12. Hansman GS, Katayama K, Maneekarn N, *et al.* (2004) Genetic diversity of norovirus and sapovirus in hospitalized infants with sporadic cases of acute gastroenteritis in Chiang Mai, Thailand. *J Clin Microbiol* **42**: 1305–1307.
13. Katayama K, Miyoshi T, Uchino K, *et al.* (2004) Novel recombinant sapovirus. *Emerg Infect Dis* **10**: 1874–1876.
14. Katayama K, Shirato-Horikoshi H, Kojima S, *et al.* (2002) Phylogenetic analysis of the complete genome of 18 Norwalk-like viruses. *Virology* **299**: 225–239.

15. Lochridge VP, Hardy ME. (2003) Snow Mountain virus genome sequence and virus-like particle assembly. *Virus Genes* **26**: 71–82.
16. Vinje J, Green J, Lewis DC, *et al.* (2000) Genetic polymorphism across regions of the three open reading frames of "Norwalk-like viruses". *Arch Virol* **145**: 223–241.
17. Jiang X, Espul C, Zhong WM, *et al.* (1999) Characterization of a novel human calicivirus that may be a naturally occurring recombinant. *Arch Virol* **144**: 2377–2387.
18. Duizer E, Schwab KJ, Neill FH, *et al.* (2004) Laboratory efforts to cultivate noroviruses. *J Gen Virol* **85**: 79–87.
19. Baric RS, Yount B, Lindesmith L, *et al.* (2002) Expression and self-assembly of norwalk virus capsid protein from Venezuelan equine encephalitis virus replicons. *J Virol* **76**: 3023–3030.
20. Harrington PR, Lindesmith L, Yount B, *et al.* (2002) Binding of Norwalk virus-like particles to ABH histo-blood group antigens is blocked by antisera from infected human volunteers or experimentally vaccinated mice. *J Virol* **76**: 12335–12343.
21. Harrington PR, Yount B, Johnston RE, *et al.* (2002) Systemic, mucosal, and heterotypic immune induction in mice inoculated with Venezuelan equine encephalitis replicons expressing Norwalk virus-like particles. *J Virol* **76**: 730–742.
22. Jiang X, Wang M, Graham DY, Estes MK. (1992) Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *J Virol* **66**: 6527–6532.
23. Prasad BV, Hardy ME, Jiang X, Estes MK. (1996) Structure of Norwalk virus. *Arch Virol Suppl* **12**: 237–242.
24. Taube S, Kurth A, Schreier E. (2005) Generation of recombinant Norovirus-like particles (VLP) in the human endothelial kidney cell line 293T. *Arch Virol* **150**: 1425–1431.
25. Barcena J, Verdaguer N, Roca R, *et al.* (2004) The coat protein of Rabbit hemorrhagic disease virus contains a molecular switch at the N-terminal region facing the inner surface of the capsid. *Virology* **322**: 118–134.
26. Bertolotti-Ciarlet A, White LJ, Chen R, *et al.* (2002) Structural requirements for the assembly of Norwalk virus-like particles. *J Virol* **76**: 4044–4055.
27. Prasad BV, Rothnagel R, Jiang X, Estes MK. (1994) Three-dimensional structure of baculovirus-expressed Norwalk virus capsids. *J Virol* **68**: 5117–5125.

28. Prasad BV, Hardy ME, Dokland T, *et al.* (1999) X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**: 287–290.
29. Chen R, Neill JD, Noel JS, *et al.* (2004) Inter- and intragenus structural variations in caliciviruses and their functional implications. *J Virol* **78**: 6469–6479.
30. Prasad BV, Matson DO, Smith AW. (1994) Three-dimensional structure of calicivirus. *J Mol Biol* **240**: 256–264.
31. Thouvenin E, Laurent S, Madelaine MF, *et al.* (1997) Bivalent binding of a neutralising antibody to a calicivirus involves the torsional flexibility of the antibody hinge. *J Mol Biol* **270**: 238–246.
32. Hansman GS, Matsubara N, Oka T, *et al.* (2005) Deletion analysis of the sapovirus VP1 gene for the assembly of virus-like particles. *Arch Virol* **150**: 2529–2538.
33. Lochridge VP, Jutila KL, Graff JW, Hardy ME. (2005) Epitopes in the P2 domain of norovirus VP1 recognized by monoclonal antibodies that block cell interactions. *J Gen Virol* **86**: 2799–2806.
34. White LJ, Ball JM, Hardy ME, *et al.* (1996) Attachment and entry of recombinant Norwalk virus capsids to cultured human and animal cell lines. *J Virol* **70**: 6589–6597.
35. Hansman GS, Natori K, Shirato-Horikoshi H, *et al.* (2006) Genetic and antigenic diversity among noroviruses. *J Gen Virol* **87**: 909–919.
36. Gray JJ, Jiang X, Morgan-Capner P, *et al.* (1993) Prevalence of antibodies to Norwalk virus in England: detection by enzyme-linked immunosorbent assay using baculovirus-expressed Norwalk virus capsid antigen. *J Clin Microbiol* **31**: 1022–1025.
37. Jiang X, Cubitt D, Hu J, *et al.* (1995) Development of an ELISA to detect MX virus, a human calicivirus in the snow Mountain agent genogroup. *J Gen Virol* **76**: 2739–2747.
38. Jiang X, Wang J, Estes MK. (1995) Characterization of SRSVs using RT-PCR and a new antigen ELISA. *Arch Virol* **140**: 363–374.
39. Kobayashi S, Sakae K, Natori K, *et al.* (2000) Serotype-specific antigen ELISA for detection of Chiba virus in stools. *J Med Virol* **62**: 233–238.
40. Dingle KE. (2004) Mutation in a Lordsdale norovirus epidemic strain as a potential indicator of transmission routes. *J Clin Microbiol* **42**: 3950–3957.
41. Noel JS, Fankhauser RL, Ando T, *et al.* (1999) Identification of a distinct common strain of "Norwalk-like viruses" having a global distribution. *J Infect Dis* **179**: 1334–1344.

42. Nilsson M, Hedlund KO, Thorhagen M, *et al.* (2003) Evolution of human calicivirus RNA *in vivo*: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. *J Virol* **77**: 13117–13124.
43. Green KY, Ando T, Balayan MS, *et al.* (2000) Taxonomy of the caliciviruses. *J Infect Dis* **181 Suppl 2**: S322–S330.
44. Konig M, Thiel HJ, Meyers G. (1998) Detection of viral proteins after infection of cultured hepatocytes with rabbit hemorrhagic disease virus. *J Virol* **72**: 4492–4497.
45. Martin Alonso JM, Casais R, Boga JA, Parra F. (1996) Processing of rabbit hemorrhagic disease virus polyprotein. *J Virol* **70**: 1261–1265.
46. Meyers G, Wirblich C, Thiel HJ, Thumfart JO. (2000) Rabbit hemorrhagic disease virus: genome organization and polyprotein processing of a calicivirus studied after transient expression of cDNA constructs. *Virology* **276**: 349–363.
47. Wirblich C, Sibilia M, Boniotti MB, *et al.* (1995) 3C-like protease of rabbit hemorrhagic disease virus: identification of cleavage sites in the ORF1 polyprotein and analysis of cleavage specificity. *J Virol* **69**: 7159–7168.
48. Wirblich C, Thiel HJ, Meyers G. (1996) Genetic map of the calicivirus rabbit hemorrhagic disease virus as deduced from *in vitro* translation studies. *J Virol* **70**: 7974–7983.
49. Sosnovtsev SV, Garfield M, Green KY. (2002) Processing map and essential cleavage sites of the nonstructural polyprotein encoded by ORF1 of the feline calicivirus genome. *J Virol* **76**: 7060–7072.
50. Sosnovtsev SV, Sosnovtseva SA, Green KY. (1998) Cleavage of the feline calicivirus capsid precursor is mediated by a virus-encoded proteinase. *J Virol* **72**: 3051–3059.
51. Sosnovtseva SA, Sosnovtsev SV, Green KY. (1999) Mapping of the feline calicivirus proteinase responsible for autocatalytic processing of the nonstructural polyprotein and identification of a stable proteinase-polymerase precursor protein. *J Virol* **73**: 6626–6633.
52. Belliot G, Sosnovtsev SV, Mitra T, *et al.* (2003) *In vitro* proteolytic processing of the MD145 norovirus ORF1 nonstructural polyprotein yields stable precursors and products similar to those detected in calicivirus-infected cells. *J Virol* **77**: 10957–10974.
53. Blakeney SJ, Cahill A, Reilly PA. (2003) Processing of Norwalk virus nonstructural proteins by a 3C-like cysteine proteinase. *Virology* **308**: 216–224.

54. Hardy ME, Crone TJ, Brower JE, Ettayebi K. (2002) Substrate specificity of the Norwalk virus 3C-like proteinase. *Virus Res* **89**: 29–39.
55. Liu B, Clarke IN, Lambden PR. (1996) Polyprotein processing in Southampton virus: identification of 3C-like protease cleavage sites by *in vitro* mutagenesis. *J Virol* **70**: 2605–2610.
56. Liu BL, Viljoen GJ, Clarke IN, Lambden PR. (1999) Identification of further proteolytic cleavage sites in the Southampton calicivirus polyprotein by expression of the viral protease in E. coli. *J Gen Virol* **80**: 291–296.
57. Seah EL, Marshall JA, Wright PJ. (1999) Open reading frame 1 of the Norwalk-like virus Camberwell: completion of sequence and expression in mammalian cells. *J Virol* **73**: 10531–10535.
58. Seah EL, Marshall JA, Wright PJ. (2003) Trans activity of the norovirus Camberwell proteinase and cleavage of the N-terminal protein encoded by ORF1. *J Virol* **77**: 7150–7155.
59. Someya Y, Takeda N, Miyamura T. (2000) Complete nucleotide sequence of the chiba virus genome and functional expression of the 3C-like protease in Escherichia coli. *Virology* **278**: 490–500.
60. Oka T, Katayama K, Ogawa S, *et al.* (2005) Proteolytic processing of sapovirus ORF1 polyprotein. *J Virol* **79**: 7283–7290.
61. Herbert TP, Brierley I, Brown TD. (1997) Identification of a protein linked to the genomic and subgenomic mRNAs of feline calicivirus and its role in translation. *J Gen Virol* **78**: 1033–1040.
62. Sosnovtsev S, Green KY. (1995) RNA transcripts derived from a cloned full-length copy of the feline calicivirus genome do not require VpG for infectivity. *Virology* **210**: 383–390.
63. Daughenbaugh KF, Fraser CS, Hershey JW, Hardy ME. (2003) The genome-linked protein VPg of the Norwalk virus binds eIF3, suggesting its role in translation initiation complex recruitment. *Embo J* **22**: 2852–2859.
64. Meyers G, Wirblich C, Thiel HJ. (1991) Genomic and subgenomic RNAs of rabbit hemorrhagic disease virus are both protein-linked and packaged into particles. *Virology* **184**: 677–686.
65. Katayama K, Shirato-Horikoshi H, Kojima S, *et al.* (2002) Phylogenetic analysis of the complete genome of 18 Norwalk-like viruses. *Virology* **299**: 225–239.
66. Asanaka M, Atmar RL, Ruvolo V, *et al.* (2005) Replication and packaging of Norwalk virus RNA in cultured mammalian cells. *Proc Natl Acad Sci USA* **102**: 10327–10332.

67. Katayama K, Hansman GS, Oka T, *et al.* (2006) Investigation of norovirus replication in a human cell line. *Arch Virol* **151**: 1291–1300.
68. Clarke IN, Lambden PR. (2000) Organization and expression of calicivirus genes. *J Infect Dis* **181 Suppl 2**: S309–S316.
69. Glass PJ, White LJ, Ball JM, *et al.* (2000) Norwalk virus open reading frame 3 encodes a minor structural protein. *J Virol* **74**: 6581–6591.
70. Wobus CE, Karst SM, Thackray LB, *et al.* (2004) Replication of Norovirus in cell culture reveals a tropism for dendritic cells and macrophages. *PLoS Biol* **2**: e432.
71. McGuffin LJ, Bryson K, Jones DT. (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16**: 404–405.

This page intentionally left blank

Chapter 15

Mathematical Approaches for Stoichiometric Quantification in Studies of Viral Assembly and DNA Packaging

Peixuan Guo[*,†], *Jeremy Hall*[†] *and Tae Jin Lee*[†]

Viral assembly involves multi-step reactions. It is relatively easy to determine the concentration of the initiating substrates and the final products. However, due to the rapidity of the reaction, it is very difficult to isolate and characterize or to elucidate intermediate states of the reaction. Stoichiometry determination is critical to the understanding of the mechanism in viral assembly. It would therefore be desirable to have a reliable method to determine the stoichiometry and the absolute copy number of the substrates or enzymes that actively participate in multi-step reactions.

The emergent field of nanotechnology generally involves the characterization, manipulation, modification, and/or assembly of organized materials on the nanoscale level. The size of viruses is between tens to hundreds of nanometers. This range falls into the dimension that is the focus of study in nanotechnology. Stoichiometry quantification will facilitate the application of viral materials in nanoscience and nanotechnology. Viruses contain a wide variety of nanomachines and ordered structures, including motors, arrays, pentons, and hexons. The novelty and ingenious design of viral particles and

*Corresponding author, 1438 Vontz Center for Molecular Studies, University of Cincinnati, Cincinnati, OH 45267, USA; Phone: (513) 558-0041; Fax: (513) 558-0024; Email: guopn@ucmail.uc.edu

†Department of Biomedical Engineering, College of Engineering and College of Medicine, University of Cincinnati, OH 45267, USA.

structural components have long inspired the development of biomimetics for nanodevices. One of the current hot topics in viral research is to make these machines as viable and effective as possible outside infectious viral particles. Viral structure components and assembly intermediates are exciting building blocks in nanotechnological and bionanotechnological applications. In general, viral structures are typically formed by multimers of gene products. In order to apply such viral structures in nanotechnology, knowledge of the stoichiometry of biological components is essential. Approaches in traditional nanotechnology are often distinguished between "top down" and "bottom up" approaches. In virology, functional nanostructures have been studied by both "bottom up" and "top down" approaches for decades, despite not using nanotechnological terminology to describe the process. Both in "top down" and "bottom up" approaches, stoichiometric determination of the components of the nanostructure is critical.

Although the absolute concentration of the substrates and the intermediate is difficult to be determined, relative parameters, such as dilution factors, percentage, and probability, as well as shapes and slopes of titration curves, can be utilized to facilitate a more accurate determination. A variety of biological systems and molecular processes have been successfully analyzed by mathematical methods. This review will only focus on the mathematical methods employed, utilized, or developed in the author's laboratory.

Introduction

Stoichiometry Quantification to Facilitate the Understanding on the Mechanism in Viral Assembly and DNA Packaging

Viruses are composed of both proteins and nucleic acids, either DNA or RNA, that serve as the genetic material. Their genome is enclosed in the protein shell or capsid. After the synthesis of structural proteins and the viral genome, these components must interact with one another to form a complete virion, a process referred to as virus assembly.[1–5] One striking feature in the assembly of many linear double-stranded DNA (dsDNA) viruses—including adenovirus,[6–10] herpes virus,[11–13] pox virus,[14–16] and bacteriophages T1,[17] T3,[18] T4,[19] T5,[20] T7,[21,22] P1,[23] P2,[24] P4,[25] P22,[26–28] Mu,[29] phi21,[30] phi29,[31–35] SPO1,[36,37] SPP1,[38,39] lambda,[40–43] and their relatives[44] — is that the

viral genome is packaged swiftly and processively (~140 bp/s at 25°C) into the viral prohead to a concentration characteristic of DNA crystals (~500 mg/ml) during maturation.[45–47]

Many biological processes involve multiple components in multistep reactions. Progress in the quantitative analysis of these systems has been hampered by a lack of viable methods. It would therefore be desirable to have a reliable method to determine the stoichiometry of the substrates or enzymes that actively participate in intermediate reactions. Although the exact concentrations of the substrates and products of intermediate reactions are difficult to determine, relative parameters, such as dilution factors, percentage of mutants, and probability, as well as shapes and slopes of dilution curves, can be more accurately determined. A variety of biological systems and molecular processes have been successfully analyzed by mathematical methods.[48–54]

Most viral assembly processes involve single-assembly pathways.[2,4,5,55] The applications of mathematical approaches to determine the stoichiometry of viral structural components are relatively simple and reliable. With careful assessment of the substrate concentration of the structural components, the approaches would be applicable to both *in vitro* and *in vivo* assembly systems. For *in vitro* assembly, the determination of substrate concentration is very straightforward. For *in vivo* studies, the concentration of the structural components can be manipulated by controlling the strength of the promoter and can be determined by SDS gels, proteomics, or other available quantitative approaches.

Stoichiometry Quantification to Facilitate the Application of Viral Materials in Nanoscience and Nanotechnology

The emergent field of nanotechnology generally involves the characterization, manipulation, modification, and/or assembly of organized materials on the nanoscale level,[56,57] thereby helping to form supramolecular structures.[58–60] These materials can then be used as building blocks for the construction of larger devices and systems.

Viral components hold great potential for nanotechnology because viruses, due to their genome size and lifecycle restrictions, are typically required to assemble functional structures efficiently from a limited number of components, following a comparatively simple construction plan.[2,3,5,61] The size of viruses is between tens to hundreds of nanometers. This range falls into the dimension that is the focus of study in nanotechnology. Viruses contain a wide variety of nanomachines and ordered structures,[62] including motors,[47,63–69] arrays,[70–73] pentons, and hexons. The novelty and ingenious design of viral particles and structural components have long inspired the development of biomimetics for nanodevices.[56,59] Structural components and intermediates of assembly in viruses are exciting building blocks in nanotechnological and bionanotechnological applications. One of the current hot topics in viral research is to make these machines as viable and effective as possible outside infectious viral particles.[74] The application of these structures and their derivatives includes the detection of pathogens, the delivery of drugs,[75] and the therapy and diagnosis of diseases.[76] Additional applications include the gearing of other nanodevices; the driving of molecular sorters; the building of intricate arrays and chips; and the operation of new electronic and optical devices,[77] including nanoelectromechanical systems (NEMS)[78] and molecular sensors or complex actuators.[79]

In general, viral structures are typically formed by multimers of gene products. In order to apply such viral structures in nanotechnology, knowledge of the stoichiometry of biological components is essential. Many physical and optical approaches have been successfully used in viral stoichiometry quantification, including the use of STEM.[80–83]

Readers are referred to these excellent review articles for detail. Although many successful cases have been reported in the use of these techniques for stoichiometry determination, direct quantification of viral components and nanoparticles has been tedious and limited in many cases. Since biological materials are often too soft and too small for atomic force microscopy, it has been difficult to yield sufficient resolution in negatively stained preparations, and to provide enough contrast in (cryo) electron microscopy due to low electron density, especially for nucleic acids.

Approaches in traditional nanotechnology are often distinguished between "top down" and "bottom up" approaches.[56,84,85] However, in virology, functional nanostructures have been studied by both "bottom up" and "top down" approaches for decades,[55,86–89] despite not using these specific terms to describe the process. "Top down" approaches include structural studies on viral components or the dissection of the complete viral particles into subunits or single molecules, using various methods of molecular biology.[8,86,90–93] "Bottom up" approaches exploit the efficiency of viral assembly. Not only viral components, but also complete infectious virions, can be assembled *in vitro* from single molecules or synthetic materials.[31,35,51,52,87,88,94–98] In both "top down" and "bottom up" approaches, stoichiometric determination of the components of the nanostructure is critical.

Hill Cooperativity Coefficient

The use of mathematical rationales in combination with experimental analyses is a versatile and powerful tool for stoichiometric quantification. Biological systems and molecular processes have been analyzed by a variety of mathematical methods, depending on the desired scientific objective.[48–52,54]. One method widely used to analyze binding equilibrium in ligand-receptor interaction is the application of the Hill equation. The most common measure of cooperativity was produced by A. V. Hill, when he devised a method based on plotting the partial pressure of oxygen against the fractional saturation of myoglobin and hemoglobin.[99] He defined the steepness of the slope at 50% saturation in a double-logarithmic graph to be the Hill coefficient. If this coefficient is 1, representing only one binding site, this indicates complete non-cooperativity, as was found for myoglobin. With only one binding site, it would not be possible for there to be any interaction between two myoglobin molecules. Hill coefficients larger than one indicate positive cooperativity, as is the case with hemoglobin, where the value is close to 3. An adaptation of the basic equation from which the Hill coefficient is derived is: $n = d \log[\Upsilon/(1 - \Upsilon)]/d \log X$, where Υ is the overall fractional saturation, and X is the ligand activity. This equation provides a method not only to measure the affinity of a ligand for a receptor, but also to estimate

the number of ligand molecules needed for receptor binding in order to execute appropriate functions.[100–103] The number of ligands can only be estimated by the Hill coefficient under conditions where extreme positive cooperativeness occurs. That is, the binding of the second molecule is dependent on the first one. For this reason, the Hill coefficient is only an indication of the interaction coefficient reflecting cooperativeness, rather than in itself being a reliable method for stoichiometry determination.[104] In addition, in a simple sequential or an independent binding reaction with up to ten binding sites for neutral cooperativeness, the Hill coefficient is always less than 2.[105] To this end, it would be desirable to have a more reliable and simple method to determine the stoichiometry of structural components or enzymes that actively participate in intermediate reactions. Nevertheless, in conjunction with other approaches, the Hill coefficient can be useful in further corroborating other methods to determine stoichiometry.

The Hill coefficient has been applied to the study of pRNA stoichiometry of the DNA packaging motor of phi29.[106] If dimers are the building blocks, there are two possible pathways to assemble a hexamer. The first one is the independent addition of individual dimers ($2 \times 3 = 6$). The second way is the cooperative addition of dimers through hand-in-hand interaction to recruit the incoming dimers ($2 + 2 + 2 = 6$). Binding analysis reveals that the Hill coefficient[105] is 2.5, strongly suggesting that on the connector there are three binding sites for dimer binding and that the binding process is cooperative. Therefore, the sequence of hexamer formation is predicted to be ($2 + 2 + 2 = 6$),[106] and the usefulness of the Hill coefficient to support stoichiometric conclusions is demonstrated.

The use of Binomial Distribution for the Determination of Stoichiometry

Principle

Binomial distribution has often been used as an effective means for stoichiometric determination.[107,108] This is particularly applicable to biological components containing two functional domains, one of

which is responsible for the binding of the component to a specific target, and the other for an essential biological activity other than binding. A lethal mutant of this component could be constructed by introducing a mutation into the functional domain while keeping the binding domain intact. If this mutant component retains a binding affinity equal to the wild-type component, then the stoichiometry of this component could be determined by mixing the wild type with the mutant in the reaction to access the probability of the distribution and combination of the mutant and wild type. This method is intended to determine the stoichiometry of components that participate in one single step of the reaction.

Here "receptor" represents the reaction assemblage, and "ligand" represents the component to be measured for stoichiometry. The probability that the receptor possesses a certain amount of mutant and a certain amount of wild-type ligand can be predicted through binomial expansion (Eq. (1)).

$$(p+q)^Z = \binom{Z}{0}p^Z + \binom{Z}{1}p^{Z-1}q + \binom{Z}{2}p^{Z-2}q^2 + \binom{Z}{Z-1}pq^{Z-1} + \binom{Z}{Z}q^Z = \sum_{M=0}^{Z}\binom{Z}{M}p^{Z-M}q^M. \quad (1)$$

where p and q are the known ratios (percentage) of mutant and wild-type ligands, respectively, in the reaction mixture, where $p + q = 1$; Z is the total number of ligands per receptor, that is, the unknown stoichiometry to be determined; M is the number of mutants bound to the receptor. Therefore:

$$\binom{Z}{M} = \frac{Z!}{M!(Z-M)!}\ . \quad (2)$$

The probability of different combinations of mutant and wild type can be predicted to produce theoretical curves. Using various ratios of mutant to wild-type components in an experimental setting, the percentage of mutant vs. the yield of products of the empirical data can

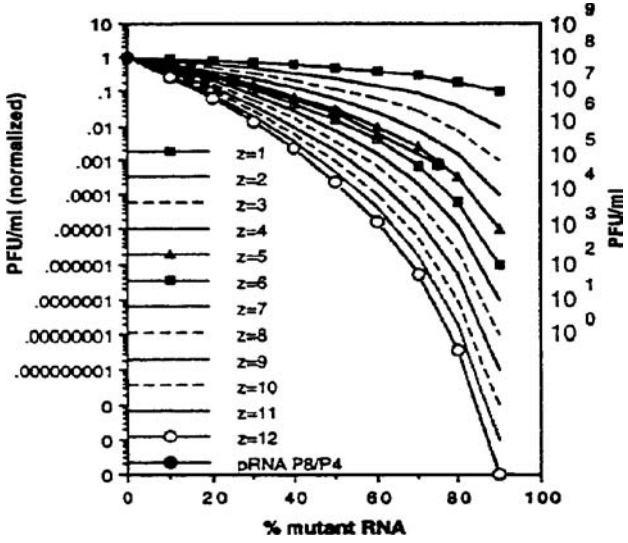


Fig. 1. Theoretical and empirical plots of % mutant RNA yield of infectious virions in *in vitro* assembly assays. *Z* is the total pRNA number per procapsid varied from 1 to 12, supposing that the minimal number of the bound mutant pRNA required to block DNA packaging is 1. (Adapted from Ref. 107, with permission from the author and the publisher, American Society for Microbiology.)

be plotted and compared to a series of predicted curves to find a best fit (Fig. 1). In this method, there will be two unknown parameters. One is the stoichiometry *Z*, and the other is the copy number, *X*, of mutant that is needed to block the reaction. In the real experiment, the productive reaction is "dominant" and can be observed, while the abortive reaction is "recessive" and can only be predicted, but not observed. Therefore, the equation can be used to determine two parameters in separate experiments:

Case one assumes that the stoichiometry, *Z*, is unknown, but the number, *X*, that is sufficient to block the reaction is one. A series of curves can then be generated; for example, several curves can be generated where *Z* varies from 1 to 30 (Fig. 2). Curves of observed empirical data can then be superimposed onto the predicted curves to find a match, and thus the stoichiometry can be deduced from the

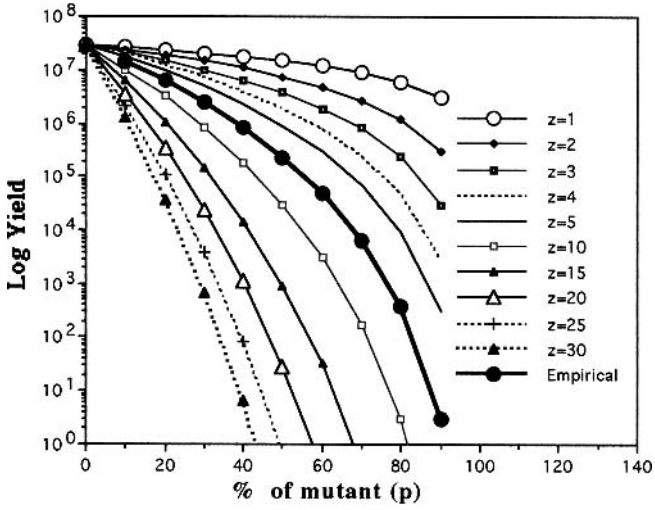


Fig. 2. Theoretical and empirical plot of % mutant the log of yield. Predictions were made with Eq. (1), supposing that the stoichiometry, *Z*, varies 1 to 30, and the minimal number, *X*, of the bound mutant required to block the activity is 1.

overlapping curves. When this is combined with Method 2, which utilizes log/log plotting, the stoichiometry can be determined.

Case two uses components with known stoichiometry or assumes that *Z* is known but the number, *X*, of mutants required to block the reaction is unknown. A series of curves will be produced and used for comparison with empirical data (Fig. 3). It is crucial to use a mutant component that can inhibit packaging completely without cooperative or partial activity.

One potential avenue of research is to explore whether previous approaches and developed equations for RNA stoichiometric determination using binomial distribution could be applied to protein components. This would lead to a conclusion as to whether or not these methods could have general application. To test the formula of binomial distribution for stoichiometric determination of proteins, it is important to utilize a protein with well-elucidated structure, clearly defined domains, unambiguous stoichiometry, and reliable methods of biological assay.

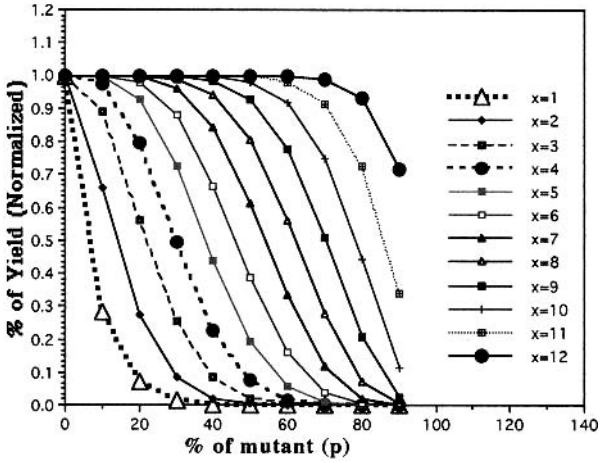


Fig. 3. Theoretical plot of % mutant the yield. Predictions were made with Eq. (1), supposing that the total stoichiometry, *Z*, is equal to 12, and the minimal number, *X*, of the bound mutant required to block the activity is 1 to 12, respectively.

Linear dsDNA viruses package their genome into a preformed procapsid.[3,5,55] The packaging motor involves a dodecameric protein connector[109–112] and two nonstructural components with certain characteristics typical of ATPase.[113] In bacterial virus phi29, these two nonstructural components are gp16 (a protein)[113,114] and pRNA.[115] Both of them have ATPase activity required for DNA encapsidation,[113,116–118] and a small (120-base) viral RNA (pRNA) also plays an essential role in the process.[115,119] Phylogenetic analyses of the virus-encoded pRNAs[115,119] from other phages show similar secondary structures.[120,121] Mg^{++} induces a refolding of pRNA. This conformational change confers the pRNA to bind to the portal vertex (the site on procapsids where DNA packaging occurs) of procapsids.[122] Approximately six pRNAs are attached to procapsids purified from infected cells,[123] although the number required for DNA packaging is unknown. The pRNA appears to leave the procapsid after DNA packaging is completed (i.e. before tail assembly)[124] and accomplish the DNA translocation process by their sequential action (Fig. 4).[125] pRNA contains two functional domains (Fig. 5). The procapsid

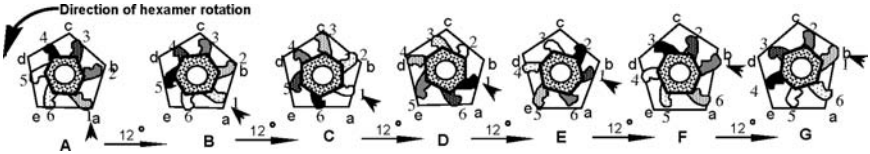


Fig. 4. A model depicting sequential action of six pRNAs to gear the DNA translocation machine. Hexagon → connector; Surrounding pentagon → capsid membrane; six protrusions → six pRNAs. Arrows point to the different energetic states of pRNA 1 with variable patterns to illustrate different energetic conformation. A to G → six steps of 12° rotation, (a 5/6-fold mismatch generates 30 equivalent orientations, 360° ÷ 30 = 12°). Each pRNA rotates 72° after six steps to move from one vertex to an adjacent vertex (pRNA 1 moves from vertex a to b). If one ATP is needed for one pRNA to move from one vertex to another via one cycle of contraction and relaxation, then 30 ATPs (6 RNA X 5 vertex) are needed for a 360^0 rotation that cause the translocation of one helical turn or 10.5 bases of DNA. (Adapted from Ref. 124, with permission from the author and the publisher, American Society for Microbiology.)

binding domain is located in the central region,[106,123,126,127] bases 23–97, while the 5'/3'-paired ends are required for an unknown role involved in DNA translocation into the procapsid.[123,127–130] Mutant pRNAs with mutations in the 5'/3'-paired ends can compete with wild-type pRNA for procapsid binding and inhibit phage assembly *in vitro* and *in vivo*.[107,131]

It is able to assemble infectious phi29 virions *in vitro* using nine purified and two partially purified proteins overproduced from cloned genes as well as synthetic pRNA and phi29 genomic DNA (Fig. 6).[132] Mixing purified procapsids, pRNA, packaging enzyme gp16, genomic DNA-gp3 (gp3 is a viral protein gp3 covalently attached to each 5' end), ATP, and Mg^{++} results in the accumulation of DNA-filled heads.[35] The subsequent addition of purified phage neck and tail proteins converts the DNA-filled heads into infectious virions (Fig. 7).[31,132] Omission of any one of the essential phi29 assembly components from the reaction results in no plaque formation,[31,35,132] thus providing an assay with a sensitivity of over eight orders of magnitude. The system is sensitive enough that pRNAs with 10^5- or 10^6-fold reductions in DNA

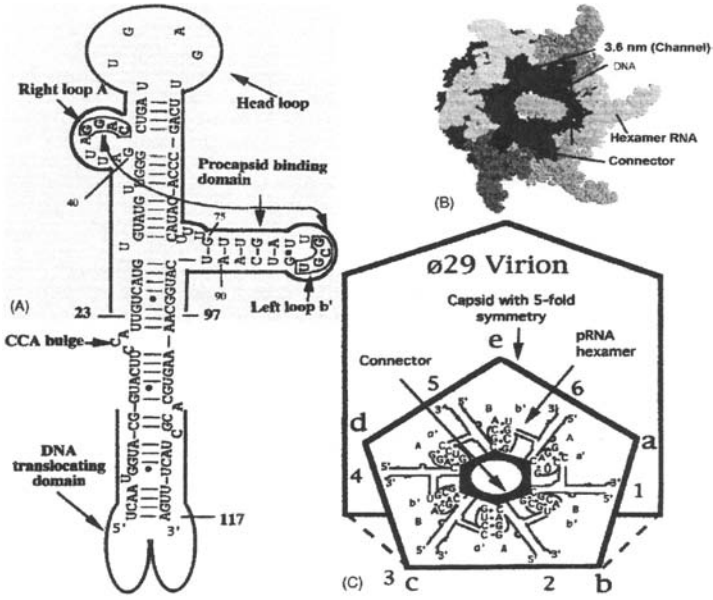


Fig. 5. Secondary structure of pRNA on phi29 bacteriophage. (A) Secondary structure of wild-type pRNA *A-b′*. The binding domain (shaded area) and the DNA translocation domain are marked with bold lines. The four bases in the right and left loops, which are responsible for inter-RNA interactions, are boxed. (B) A 3D-model of the viral packaging motor. (C) Diagrams depicting the formation of pRNA hexamer ring and its location in the phi29 DNA packaging machine. Hexamer formation is via right-and-left-loop sequence interactions of pRNA *A-b′* and *B-a′*. (Adpated from Ref. 151, with permission from the author and the publisher, Elsevier.)

packaging activity can be detected, which was not possible with previously used DNA packaging assays.

Therefore, it is now an appropriate time to test whether the binomial distribution approach can be applied to the determination of protein stoichiometry.

Determination of the Stoichiometry, Z

Equation (1) can be used to predict the level of mutants. For example, if the total number of pRNAs per procapsid required for DNA

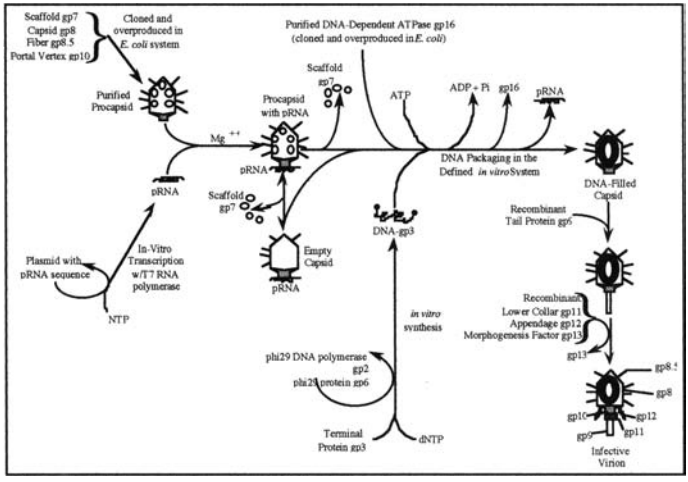


Fig. 6. Steps in the synthesis of infectious virions of pho29 *in vitro*. (Adapted from Ref. 31, with permission from the author and the publisher, American Society for Microbiology.)

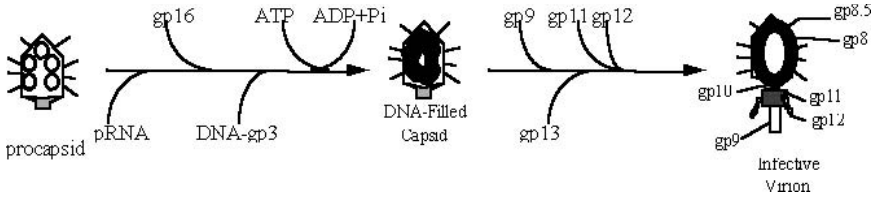


Fig. 7. Assembly pathway of phi29 *in vitro*.

packaging is three, the probability of all combinations of mutant and wild-type pRNAs on a given procapsid can be determined by the expansion of the binomial: $(p + q)^3 = p^3 + 3p^2q + 3pq^2 + q^3 = 100\%$. That is, in the procapsid population, the probability of procapsid possessing either three copies of mutant pRNA, two copies of mutant and one copy of wild type, one copy of mutant and two copies of wild type, or three copies of wild type is p^3, $3p^2q$, $+ 3pq^2$, or q^3, respectively. Suppose that there were 70% ($p \times 100\%$) mutant and 30% ($q \times 100\%$) wild-type pRNA in the reaction mixture, then, the percentage of procapsids that possessed at least two copies of wild-type pRNA would

be the sum of those possessing one copy of mutant and two copies of wild type, $3pq^2$, and those possessing three copies of wild-type mutant, q^3, that is, $3pq^2 + q^3 = 3(0.7)(0.3)2 + (0.3)3 = 0.216 = 21.6\%$. If only one mutant pRNA per procapsid was sufficient to render the procapsid unable to package DNA, then only those procapsids with three wild-type pRNAs bound (0 mutant pRNA) would package DNA.

As another example, if $Z = 6$, but only five of the six need to be wild type in order to produce final products, the relative yield will be the sum of: (1) the probability of each receptor containing five wild type; and (2) the probability of each receptor containing six wild type. The probability that a given receptor will have M mutant and W wild type bound at increasing percentages of mutant can be calculated under the conditions that $Z = 6$ and 12, respectively, using Eq. (3) (Table 1 and 2);

$$\left(\frac{Z!}{M!\,W!}\right)p^M q^W \tag{3}$$

Table 3 shows the prediction of the probability of sound reactions (products) in the presence of various ratios of mutant and wild type. In this case, it is assumed that the number of mutant, X, needed to block the reaction is one, and Z varies from one to six. That is, one mutant per single reaction is sufficient to kill the reaction, notwithstanding that the total number of said component in each single reaction is a constant, one, two, three, four, five, or six. The simplified formula q^Z was used to predict the probability of the dominant productive reaction, since, if one mutant will block the reaction, the only reaction that would be able to occur would be the probability when $W = Z$ and $M = 0$. An example of such a probability assessment, when $Z = 6$ and $X = 1$, has been shown for pRNA of phi29.[107] Since it was assumed that six copies are used and one mutant is sufficient to block, all reactions involving one, two, three, four, or five copies of mutant, respectively, would have been inactive. Only those involving six copies of wild type would be competent. The chance that a receptor will catch six copies of wild type and no mutant pRNA in a mixture containing q wild-type and p mutant pRNA is q^6.

Table 1. Probability of Receptor Possessing Various Amounts of Wild-type (N) and Mutant (M) when $Z = 6$*.

Mutant RNA (P) (%)	Probability of Procapsids With:						
	$N = 62$, $M = 0$	$N = 5$, $M = 1$	$N = 4$, $M = 2$	$N = 3$, $M = 3$	$N = 2$, $M = 4$	$N = 1$, $M = 5$	$N = 0$, $M = 6$
0	1	0	0	0	0	0	0
10	0.531441	0.354294	0.098415	0.014580	0.001215	0.0000510	0.000001
20	0.262144	0.393216	0.245760	0.081920	0.015360	0.001536	0.000064
30	0.117649	0.302526	0.324135	0.185220	0.059335	0.010206	0.000729
40	0.046656	0.186624	0.311040	0.276490	0.138240	0.036864	0.004096
50	0.015625	0.093750	0.234375	0.312500	0.234375	0.003750	0.015625
60	0.004096	0.036864	0.138240	0.276490	0.311040	0.186624	0.046656
70	0.000729	0.010206	0.059535	0.185220	0.324135	0.302526	0.117649
80	0.000064	0.001536	0.015360	0.081920	0.245760	0.393216	0.262144
90	0.000001	0.000054	0.001215	0.014500	0.098415	0.354294	0.531441
100	0	0	0	0	0	0	1

* Predictions were made various ratios of mutant (p) and wild-type (q) pRNAs. Each procapsid is assumed to contain six copies of pRNA. N and M are number of copies of wild-type and mutant pRNAs, respectively, attached to one procapsid. $M + N = Z = 6$; $p + q = 1$. (Adapted from Ref. 107, with permission from the author and the publisher, American Society for Microbiology.)

Table 2. Probability of Receptor Possessing Various Amounts of Wild-type (*N*) and Mutant (*M*) when *Z* = 12*

	Probability of Receptor with the Combination of:												
%Mutant (*P*)	*W* = 12, *M* = 0	*W* = 11, *M* = 1	*W* = 10, *M* = 2	*W* = 9, *M* = 3	*W* = 8, *M* = 4	*W* = 7, *M* = 5	*W* = 6, *M* = 6	*W* = 5, *M* = 7	*W* = 4, *M* = 8	*W* = 3, *M* = 9	*W* = 2, *M* = 10	*W* = 1, *M* = 11	*W* = 0, *M* = 12
0	1	0	0	0	0	0	0	0	0	0	0	0	0
10	0.282429	0.376572	0.230127	0.085232	0.021308	0.003788	0.000491	0.000046	0.000032	0.000002	0	0	0
20	0.068719	0.206158	0.283467	0.236223	0.132875	0.053150	0.015502	0.003321	0.000519	0.000576	0.000043	0.000002	0
30	0.013841	0.071183	0.167790	0.239700	0.231139	0.158495	0.079247	0.029111	0.007797	0.001485	0.000190	0.00015	0.000005
40	0.002176	0.017414	0.063852	0.141893	0.212840	0.227030	0.176579	0.100902	0.042060	0.012457	0.002491	0.00030	0.000017
50	0.000244	0.002929	0.016113	0.053710	0.120849	0.193359	0.225585	0.193359	0.120849	0.053710	0.016113	0.00293	0.00024
60	0.000017	0.000301	0.002491	0.012457	0.042042	0.100902	0.176579	0.227030	0.212840	0.141893	0.063852	0.01741	0.00218
70	0.000005	0.000015	0.000190	0.001485	0.007797	0.029111	0.079247	0.158495	0.231139	0.239700	0.167790	0.07118	0.01384
80	0.000000	0.000005	0.000043	0.000057	0.000519	0.003321	0.015502	0.053150	0.132875	0.236223	0.283467	0.20616	0.06872
90	0	0	0	0.000002	0.000032	0.000046	0.000491	0.003788	0.021308	0.085232	0.230127	0.37657	0.28243
100	0	0	0	0	0	0	0	0	0	0	0	0	1

Table 3. Predicted Probability of Yields in the Presence of Various Ratios of Mutant and Wild-type Components when *Z* varies from 1 to 6 and *X* = 1*.

Mutant RNA (*P*) (%)	Probability of Φ29 Assembly with:					
	Z = 6, *X* = 0	*Z* = 5, *X* = 1	*Z* = 4, *X* = 1	*Z* = 3, *X* = 1	*Z* = 2, *X* = 1	*Z* = 1, *X* = 1
0	1	0	1	1	1	0
10	0.533333	0.600000	0.666667	0.733333	0.800000	0.9
20	0.263333	0.326667	0.400000	0.500000	0.633333	0.8
30	0.116667	0.166667	0.240000	0.333333	0.500000	0.7
40	0.046667	0.076667	0.130000	0.216667	0.366667	0.06
50	0.015667	0.031000	0.063333	0.125000	0.250000	0.5
60	0.004000	0.010000	0.027000	0.063333	0.160000	0.4
70	0.000730	0.002433	0.008000	0.027000	0.090000	0.3
80	0.000063	0.000320	0.001600	0.008000	0.040000	0.2
90	0.000001	0.000010	0.000100	0.001000	0.010000	0.1
100	0	0	0	0	0	0

* Predications were make for conditions where the number of pRNAs pre procapsid that were required for packing (*Z*) was varied from 1 to 6, while the number of mutant pRNAs pre procapsid that were sufficient to inhabit packing (*X*) was predicted to be 1. (Adapted from Ref. 107, with permission from the author and the publisher, American Society for Microbiology.)

Determination of the X, the Copy Number of Mutant Sufficient to Block Individual Reaction

DNA-packaging activity of multiple copies of pRNA may be disrupted by the incorporation of only one mutant pRNA molecule into the group.[107,131] In a series of predicted curves generated using the binomial distribution, Z was held constant at 12 while X varied from one to 12, as can be seen by the 12 separate curves (Fig. 8). The value of p varied from zero (0%) to one (100%) by increments of .1 (10%) for each predicted curve. These predicted curves assume that mutant and wild type have equal procapsid binding affinities and thus can compete equally for binding sites, and that the binding of mutant and wild type is irreversible. When X increases, the efficiency of inhibition by mutant decreases. The curve of observed empirical data could then be superimposed onto the hypothetical curves to find a match, and thus the X can be deduced from the overlapped curve.

One problem in using the binomial distribution method is the large number of theoretical curves that will be produced, and the relatively little difference between curves with similar X and Z values (i.e. $Z = 10$, $X = 1$ for one curve and $Z = 11$, $X = 1$ for another curve). However, most of the curves will be quite different from the empirical curves in shape and slope and can be immediately excluded. This will leave a relatively small number of curves to choose from. Such an approach, coupled with another method such as slopes of log/log plot curves, may produce accurate stoichiometry of pRNA.[107,125,133]

In Fig. 8, predicted curves generated using Eq. (1) under conditions of $Z = 12$ while X varies from one to 12 indicate that inhibition of phi29 as a function of % mutant pRNA becomes greater as the number of pRNAs required to inhibit DNA packaging (X) decreases. When the empirical data of inhibition of phage assembly by mutant pRNA P8/P4 were superimposed onto the predicted curves, the slope and magnitude of inhibition observed did not match any of the predicted curves. Attention should be paid to compare the slope of the curve rather than the height of the curve. Slope is influenced by the ratio of mutant and wild type, while the height of the curve is influenced by the absolute total concentration of mutant plus wild

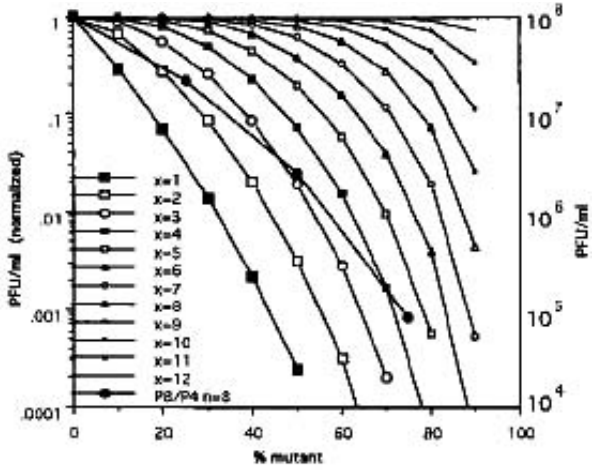


Fig. 8. Theoretical and empirical (mutant pRNA P8/P4) plot of % mutant RNA the log of yield of infectious virion in *in vitro* phage assembly assays, expressed as PFU (right) /ml or normalized (left). Predictions were made with Eq. (1), supposing that the total pRNA stoichiometry, *Z*, per procapsid is equal to 12, and the minimal number, *X*, of the bound mutant pRNA required to block DNA packaging is 1 to 12, respectively. Since we look at the slope other than the height of the curves, not a single theoretical curve could match with the empirical curve, indicating that all the hypothesized stoichiometries are incorrect. (Adapted from Ref. 107, with permission from the author and the publisher, American Society for Microbiology.)

type. The total concentration of components involved in the intermediate reaction can vary due to degradation or aggregation.

Determination of Stoichiometry by Slopes of Log/Log Plot Curves for Concentration Versus Yield

Principle

The basis for designing this method is that the slope of the curve in the log/log plot of the concentration versus the product for each component is the intrinsic parameter reflecting the stoichiometry of

the component, since the larger the stoichiometry of the component, the more dramatic the influence of the component concentration (dilution factor) on the reaction. In the plot curve, the *y*-axis is the log of the yield of the reaction, and the *x*-axis is the log of the step-up concentration, or the log of the inverse of the dilution factor starting from high dilution. The larger the stoichiometry of one component, the larger the slope of the log/log plot curve for this component will be. The slope is defined as the tangent of the angle between the curve and the *x*-axis. When plotting the curve for measuring the slope, the unit length for the *x*-axis (log of concentration) must be the same as that of the *y*-axis (log of yield).

Approaches

The dose response curves of *in vitro* viral assembly versus the concentration of various assembly components have been used as a method to estimate stoichiometry. In these dose response curves, the larger the stoichiometry of the component, the more dramatic is the influence of the dilution factor on the reaction. A slope of one indicates that one copy of the component is involved in the assembly of one virion, as is the case, for instance, in genomic DNA in phi29 assembly. A slope larger than one would indicate multiple-copy involvement (cooperative binding).

To test one parameter, all other components should be kept in excess in this assay. The relative concentration (*x*-axis, from low to high) was plotted against the yield (*y*-axis) of phi29 virions assembled (PFU/ml). The log scale was used for both the concentration and the yield, with the unit length of the log scale of the *x*-axis equal to that of the *y*-axis.[107,108,134,151] The angle between the *x*-axis and the concentration-dependent curve was measured, and the slopes (tan) of all curves were plotted against their stoichiometry. Computer programs such as Cricket Graph or Prism can be utilized to find a best-fit curve and to deduce the formula for stoichiometry determination.

The log/log plot method only considers the functional unit (oligo or complex), not the copy number of the subunits, as described

herein. Log-log plotting has revealed that the slope of the log/log concentration-dependent curve for procapsid is one, and therefore one copy of procapsid is needed for the assembly of one virion.[31,107,108,132,151]

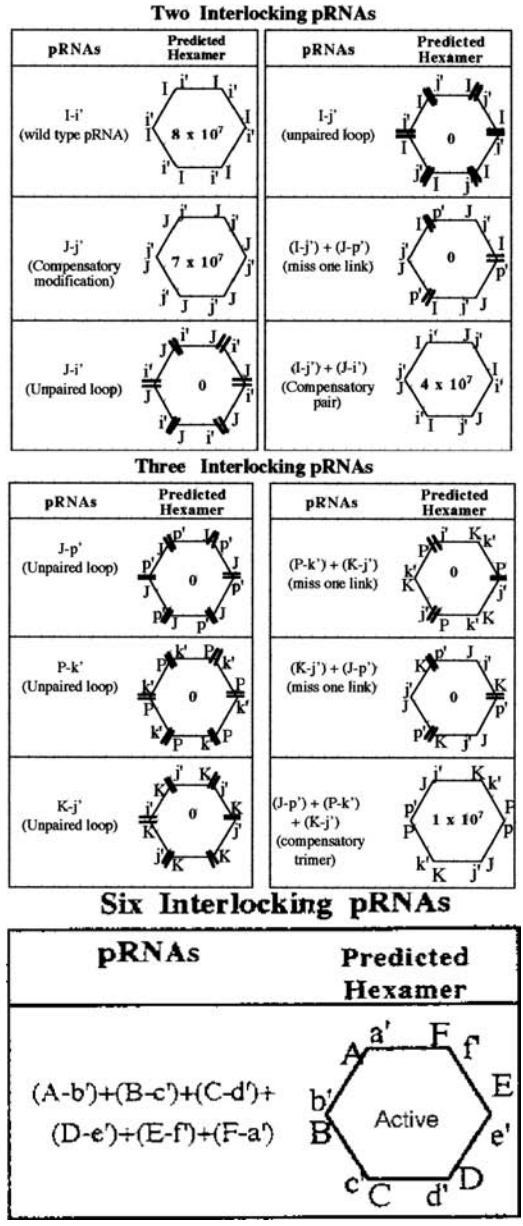
Descending of Trimeric, Dimeric, and Monomeric Slopes of Log/Log Plot Curves

There are two loops, the left- and right-hand loops,[121] in the predicted secondary structure of the pRNA. Sequences of these two loops (bases 45–48 of right-hand loop and bases 85–82 of left-hand loop) are complementary. However, extensive studies[106,108,121,122,125,135–137] have revealed that these two sequences interact intermolecularly, allowing the formation of pRNA oligomers. Several pRNAs with mutated left- and right-hand loop sequences were constructed. One pRNA building block, not the hexamer, is regarded as one functional unit. Since six monomeric pRNA are needed to build a hexamer, the stoichiometry for momomer pRNA is six. Since three pRNA dimers are needed to build a hexamer, the stoichiometry for pRNA dimer is three. Since two trimeric pRNA are needed to build a hexamer, the stoichiometry for pRNA trimer is two (Table 4). The slopes of log/log plot curves for trimeric, dimeric, and monomeric pRNA descend with the stoichiometry, suggesting that slopes and stoichiometry are correlated (Fig. 9).

Deduction of Equation for Stoichiometric Determination

The log of the concentration (*x*-axis, from low to high) can be plotted the log of the yield (*y*-axis, PFU/ml). The unit length of the log scale of *x*-axis and *y*-axis should be made equal in the plot to determine the angle between the *x*-axis and the concentration-dependent curve (Figs. 9, 10, 11). The value of tangent (slope, tan) for the angle of the procapsid, DNA-gp3, pRNA trimer, dimer, and monomer, as well as gp9, gp11, and gp12, can be obtained from experimental data (Table 5). The slopes (tan) of all curves are plotted against

Table 4. Interlocking pRNAs of Loop-Loop Interaction



(Adapted from Ref. 133, with permission from the author and the publisher, Cell Press.)

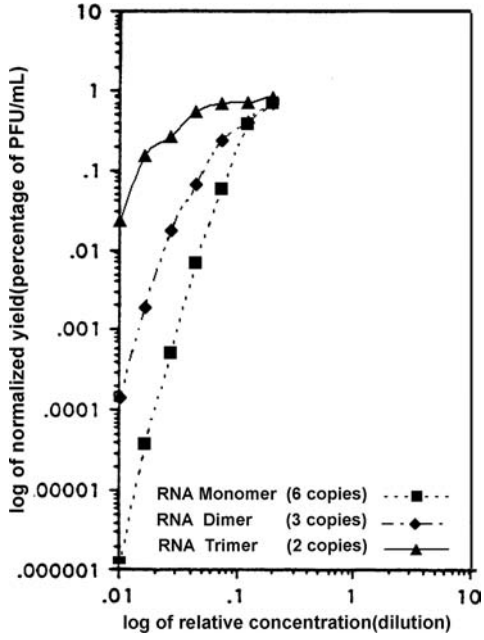


Fig. 9. Log/log plot of concentration vs. the yield of virus production for pRNA monomer, dimer, and trimer. (Adapted from Ref. 151, with permission from the author and the publisher, Elsevier.)

their stoichiometry. A computer program, such as Cricket Graph, can be used to find a best-fit curve and to deduce the formula for stoichiometric determination.

Several formulas have been tested for their suitability in the stoichiometric determination of viral assembly components in various circumstances. These formulas include polynomial ($Z = m + k_1T + k_2T^2 + \cdots$), logarithmic [$Z = (m + k^*\log(T)$], exponential ($Z = mn^{kT}$), and linear ($Z = m + kT$) equations, where Z is the stoichiometry, T is the slope that is equal to the tangent of the angle between the curve and the x-axis, and k and m are constants that are determined empirically. In the experimental data from phi29 (Table 5), the polynomial with two orders is the best fit, as suggested by the R^2 value of 0.998 (Fig. 12). When the log of dose response curves or dilution factor is plotted against the log of the

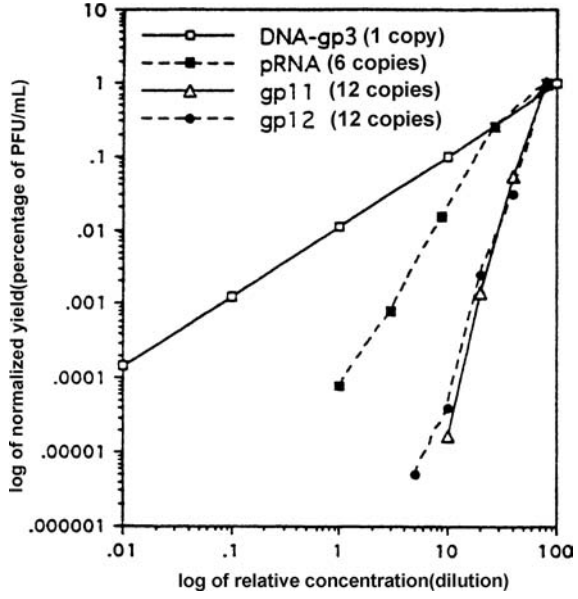


Fig. 10. Dose-response curves with log/log plot showing DNA-gp3, pRNA, gp16, gp11, and gp12 concentration dependence. The slopes of the curves for DNA are 1 (45?), suggesting that one DNA is needed for assembly of one virion. (Adapted from Ref. 151, the author and the publisher, Elsevier.)

yield of the reaction, the stoichiometry Z can be determined with an equation (Eq. (4)),

$$Z = -1.58 + 2.4193T - 0.001746T^2 \; [T \in (0, 1000), \text{ or } \angle\alpha \in (0°, 89.9°)] \tag{4}$$

where T is the slope of the curve (i.e. the tangent of ∠a), which was defined as the angle between the curve and the x-axis. This equation is also examined for its feasibility in stoichiometry determination (Tables 6 and 7). It was found that the stoichiometry could be determined by these formulas with high reliability. For example, the angle of the curve representing gp12 was determined to be 80°, so $T = \tan 80° = 5.67$. From equation $Z = -1.58 + 2.4193T - 0.001746T^2$, if

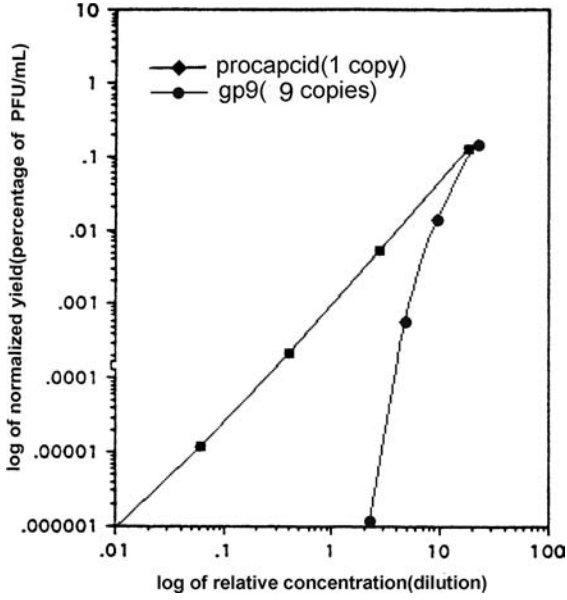


Fig. 11. Log/log plot of concentration vs. the yield of virus production for procapsid and gp9. (Adapted from Ref. 151, the author and the publisher, Elsevier.)

$T = 5.67$, $Z = 12.1$. This means that the theoretical value of the stoichiometry of gp12 is 12. This agrees with the stoichiometry of gp12 that has been determined previously.[138–141]

The issue of the reusability of components can interfere with and complicate the results when using the log/log slope assay method for stoichiometric determination. The principle for the determination of stoichiometry by this method is based on the concentration of the components. If a component is reusable, the outcome of quantification will not be reliable in comparison to components that are nonreusable. If the component is reusable, the recycled molecules will add to the pool of molecules that are able to function in *in vitro* assembly. Therefore, the concentration will appear higher than it actually is, thus the stoichiometry determined will be lower than the real number.

Table 5. Stoichiometry and Slopes for phi29 Components, such as Procapsid, DNA-gp3, pRNA Trimer, Dimer and Monomer, as well as gp9, gp11 and gp12, obtained by *In Vitro* Viral Assembly System

Components for Phi29 Virion Assembly	Function	Total Subunits/ Virion	Stoichiometry	∠ of Curve (Degrees)	Slopes (tan) Empirical	Slopes (tan) Theoretical
DNA-gp3	Genome complex	1	1	45 ± 4	1	1
Procapsid	Hold DNA	1	1	47 ± 3	1.07	1
pRNA trimer	DNA-packaging	6	2	56 ± 2	1.48	1.48
pRNA dimer	DNA-packaging	6	3	64 ± 5	2.05	2.05
pRNA monomer	DNA-packaging	6	6	72 ± 3	3.08	3.14
Gp9	Tail knob protein	10	9	77 ± 2	4.33	4.33
Gp11	Lower collar	12	12	79 ± 1	5.14	5.67
Gp12	Anti-receptor	24	12	80 ± 2	5.67	5.67

(Adapted from Ref. 151, with permission from the author and the publisher, Elsevier.)

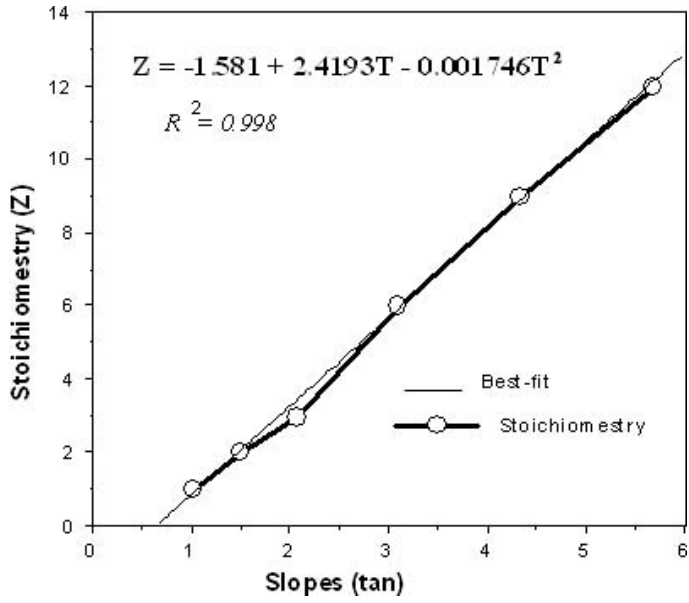


Fig. 12. Plotting of stoichiometry against slopes of dose-response log/log plot curves of components in phi29 assembly *in vitro*. The equations were derived from the best-fit curve with an *R* square of 0.998. (Adapted from Ref. 151, with permission from the author and the publisher, Elsevier.)

Common Multiples

In order to eliminate certain possible stoichiometric values for biological assemblages, attempts have been made to quantify which combinations of building blocks come together to form a larger unit, thus suggesting that these subunits make up a constituent part of the system as a whole in question. The usual process involves the manipulation of the building blocks (each of which are necessary for proper functionality of the larger subunit under investigation) to allow for complementarity, and subsequent testing to reveal the extent to which these altered building blocks cooperate with one another to form a larger subunit. A recent example of this is the study conducted on pRNA, in which two inactive mutant pRNAs with complementary loops were mixed intermolecularly and then assayed for the relative amount of activity exhibited in DNA packaging.[133,137] There was

Table 6. Theoretical Conversion Table from Stoichiometry to Slopes

Stoichiometry	*Slopes (tan)*	Stoichiometry	*Slopes (tan)*	Stoichiometry	*Slopes (tan)*
1	*1*	**9**	*4.33*	**17**	*7.72*
2	*1.48*	**10**	*4.80*	**18**	*8.14*
3	*2.05*	**11**	*5.22*	**19**	*8.56*
4	*2.31*	**12**	*5.67*	**20**	*8.98*
5	*2.73*	**13**	*6.05*	**21**	*9.40*
6	*3.14*	**14**	*6.47*	**22**	*9.82*
7	*3.56*	**15**	*6.89*	**23**	*10.24*
8	*3.97*	**16**	*7.31*	**24**	*10.66*

(Adapted from Ref. 151, with permission from the author and the publisher, Elsevier.)

already a published predicted secondary structure of the pRNA, which revealed two loops, termed the left- and right-hand loops.[121] The sequences of these two loops (specifically, bases 45–48 of the right-hand loop and bases 82–85 of the left-hand loop) are naturally complementary and were originally proposed to form a pseudoknot.[127] Further studies revealed that these two sequences interact intermolecularly, allowing the formation of pRNA oligomers. A variety of mutated left- and right-hand loop sequences for several pRNA were constructed. In the developed notation, the same letter in upper- and lower-case represented a pair of complementary sequences. For instance, in pRNA *A-a′*, the right and left loops are complementary, while for pRNA *A-b′*, the four bases in the right loop A are not complementary to the left loop *b′*. Those pRNAs with complementary loop sequences are active in DNA packaging, while those mutants with non-complementary loops are inactive.

Three sets of pRNAs were produced in order to facilitate this process.

One Set Composed of Two Engineered pRNAs: Multiple of 2

A key finding in pRNA research was that the mixing of two mutant pRNAs with trans-complementary loops restored DNA packaging

Table 7. Stoichiometry Prediction from Slopes

Degrees	Slopes	Stoichio	Degrees	Slopes	Stoichio	Degrees	Slopes	Stoichio
45.0	1.0	0.8	80.5	6.0	12.9	84.8	11.0	24.8
47.7	1.1	1.1	80.7	6.1	13.1	84.9	11.1	25.1
50.2	1.2	1.3	80.8	6.2	13.4	84.9	11.2	25.3
52.4	1.3	1.6	81.0	6.3	13.6	84.9	11.3	25.5
54.5	1.4	1.8	81.1	6.4	13.8	85.0	11.4	25.8
56.3	1.5	2.0	81.3	6.5	14.1	85.0	11.5	26.0
58.0	1.6	2.3	81.4	6.6	14.3	85.1	11.6	26.2
59.5	1.7	2.5	81.5	6.7	14.5	85.1	11.7	26.5
60.9	1.8	2.8	81.6	6.8	14.8	85.2	11.8	26.7
62.2	1.9	3.0	81.8	6.9	15.0	85.2	11.9	27.0
63.4	2.0	3.3	81.9	7.0	15.3	85.2	12.0	27.2
64.5	2.1	3.5	82.0	7.1	15.5	85.3	12.1	27.4
65.6	2.2	3.7	82.1	7.2	15.7	85.3	12.2	27.7
66.5	2.3	4.0	82.2	7.3	16.0	85.4	12.3	27.9
67.4	2.4	4.2	82.3	7.4	16.2	85.4	12.4	28.1
68.2	2.5	4.5	82.4	7.5	16.5	85.4	12.5	28.4
69.0	2.6	4.7	82.5	7.6	16.7	85.5	12.6	28.6
69.7	2.7	4.9	82.6	7.7	16.9	85.5	12.7	28.9

(Adapted from Ref. 151, with permission from the author and the publisher, Elsevier.)

(*Continued*)

Table 7. (*Continued*)

Degrees	Slopes	Stoichio	Degrees	Slopes	Stoichio	Degrees	Slopes	Stoichio
70.3	2.8	5.2	82.7	7.8	17.2	85.5	12.8	29.1
71.0	2.9	5.4	82.8	7.9	17.4	85.6	12.9	29.3
71.6	3.0	5.7	82.9	8.0	17.7	85.6	13.0	29.6
72.1	3.1	5.9	83.0	8.1	17.9	85.6	13.1	29.8
72.6	3.2	6.1	83.0	8.2	18.1	85.7	13.2	30.0
73.1	3.3	6.4	83.1	8.3	18.4	85.7	13.3	30.3
73.6	3.4	6.6	83.2	8.4	18.6	85.7	13.4	30.5
74.1	3.5	6.9	83.3	8.5	18.9	85.8	13.5	30.8
74.5	3.6	7.1	83.4	8.6	19.1	85.8	13.6	31.0
74.9	3.7	7.3	83.4	8.7	19.3	85.8	13.7	31.2
75.3	3.8	7.6	83.5	8.8	19.6	85.9	13.8	31.5
75.6	3.9	7.8	83.6	8.9	19.8	85.9	13.9	31.7
76.0	4.0	8.1	83.7	9.0	20.1	85.9	14.0	31.9
76.3	4.1	8.3	83.7	9.1	20.3	85.9	14.1	32.2
76.6	4.2	8.5	83.8	9.2	20.5	86.0	14.2	32.4
76.9	4.3	8.8	83.9	9.3	20.8	86.0	14.3	32.7
77.2	4.4	9.0	83.9	9.4	21.0	86.0	14.4	32.9

(*Continued*)

Table 7. (*Continued*)

Degrees	Slopes	Stoichio	Degrees	Slopes	Stoichio	Degrees	Slopes	Stoichio
77.5	4.5	9.3	84.0	9.5	21.2	86.1	14.5	33.1
77.7	4.6	9.5	84.1	9.6	21.5	86.1	14.6	33.4
78.0	4.7	9.8	84.1	9.7	21.7	86.1	14.7	33.6
78.2	4.8	10.0	84.2	9.8	22.0	86.1	14.8	33.8
78.5	4.9	10.2	84.2	9.9	22.2	86.2	14.9	34.1
78.7	5.0	10.5	84.3	10.0	22.4	86.2	15.0	34.3
78.9	5.1	10.7	84.3	10.1	22.7	86.2	15.1	34.6
79.1	5.2	11.0	84.4	10.2	22.9	86.2	15.2	34.8
79.3	5.3	11.2	84.5	10.3	23.2	86.3	15.3	35.0
79.5	5.4	11.4	84.5	10.4	23.4	86.3	15.4	35.3
79.7	5.5	11.7	84.6	10.5	23.6	86.3	15.5	35.5
79.9	5.6	11.9	84.6	10.6	23.9	86.3	15.6	35.7
80.0	5.7	12.2	84.7	10.7	24.1	86.4	15.7	36.0
80.2	5.8	12.4	84.7	10.8	24.3	86.4	15.8	36.2
80.4	5.9	12.6	84.8	10.9	24.6	86.4	15.9	36.4

activity. For example, pRNAs *I-j′* and *J-i′* were inactive in DNA packaging alone, but when mixed together they restored DNA packaging activity. This result can be explained by the trans-complementarily of pRNA loops, i.e. the right-hand loop *I* of pRNA *I-j′* could pair with the left-hand loop *a′* of pRNA *J-i′*. Since mixing two inactive pRNAs with interlocking loops, such as when pRNA *I-j′* and *J-i′* are mixed in a 1:1 molar ratio, resulted in production of infectious virions, the stoichiometry of the pRNA is predicted to be a multiple of two (Table 4). Together with the results from binomial distribution and serial dilution analyses, it has been confirmed that the stoichiometry of pRNA is six.

One Set Composed of Three Engineered pRNAs: Multiple of 3

Another set of mutants is composed of three pRNAs, *J-p′*, *P-k′*, and *K-j′*. This set is expected geometrically to be able to form a 3-, 6-, 9-, or 12-mer ring that carries each of the three mutants. Each individual pRNA and even mixing any two of the three mutants shows little or no activity. However, mixing all three pRNAs are with 1:1:1 ratio restores DNA packaging activity (Table 4). The lack of activity in mixtures of only two mutant pRNAs and the restored activity in mixtures of three mutant pRNAs is expected since the mutation in each pRNA are engineered in such a way that only the presence of all three pRNAs will produce a closed ring. The fact that the three inactive pRNAs are fully active when mixed together suggests that the number of pRNAs in the DNA packaging complex is a multiple of three, in addition to being a multiple of two. Thus, the number of pRNAs required for DNA packaging is a common multiple of two and three, which is six (or 12, but this number has been excluded by the approach of binomial distribution (see Section 1(a)).

One Set Composed of Six Engineered pRNAs: Multiple of 6

DNA packaging activity is also achieved by mixing six mutant pRNAs, each being inactive when used alone (Table 4). Thus, an interlocking

hexameric ring can be predicted to form by the base pairing of the interlocking loops.

The application of such a process is relatively straightforward. Mixing two mutant pRNAs with trans-complementary loops restored DNA-packaging activity.[133] For example, pRNAs *A/b′* and *B/a′* were inactive in DNA packaging alone, but when mixed together restored DNA-packaging activity; the trans-complementarity of the loops, when mixed in a 1:1 molar ratio, resulted in maximized production of infectious virions. This successful packaging strongly suggests that pRNA's stoichiometry is a multiple of two. The same trans-complementarity result was achieved when three pRNAs were combined, which suggested that the stoichiometry must be a multiple of three. Knowing that the stoichiometry must be a multiple of both two and three indicates that it must be a multiple of six, which drastically shortens the list of possible stoichiometries for this biological assemblage. Binomial distribution, as described below, is one of the corroborating methods used to confirm that six is the correct stoichiometry, to the exclusion of 12 or other possibilities.

Addition and Deletion of Monomers and Dimers to Exclude Possible Stoichiometries

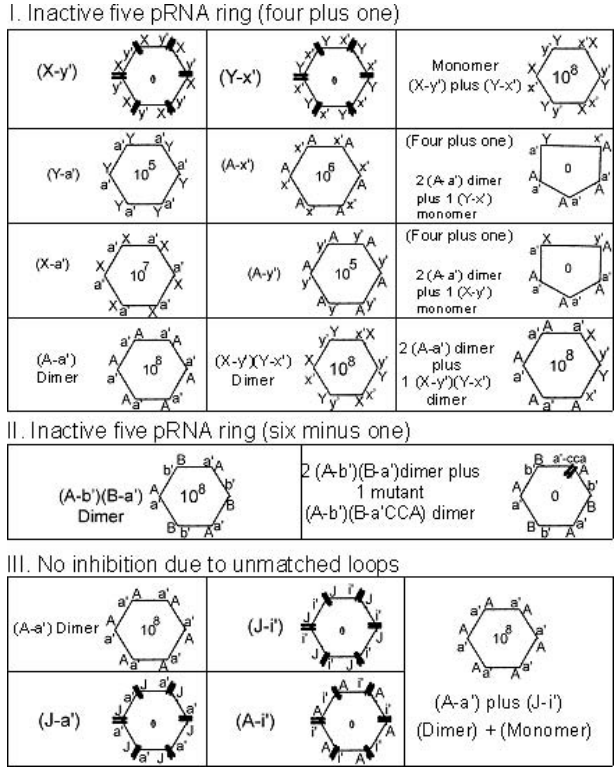
In systems where the constituent building blocks of a larger macromolecule can be manipulated, it is possible to conduct experiments that exclude or include certain values for the macromolecule's stoichiometry. In 1998, Guo and coworkers[133] and Anderson and coworkers[137] independently reported that pRNA forms a hexamer via intermolecular base-pairing between the right loop (bases 42–45) and the left loop (bases 82–85)[107,125,133,137] (see also mini-review in *Cell*[142]). Recently, there has been an ongoing, fervent debate as to whether there is a pRNA hexamer that binds to the twelve-fold symmetrical connector and serves as the rotor,[125,133,137,142–146] or a pRNA pentamer that binds to the five-fold symmetrical capsid protein to serve as the stator.[111,147,148] Cryo-EM studies on pRNA have also produced conflicting data showing that pRNA functions as a hexamer[145] or a pentamer.[111,148] As noted earlier, the stoichiometry of pRNA is the

common multiple of two and three, that is, six. This does not, however, fully disprove the possibility of the functionality of a stoichiometry of five, which has been the stoichiometry postulated in other models.[111] In the model, the five vs. six discrepancy was interpreted as that one copy of the pRNA dissociates from initial hexameric pRNAs after binding to procapsids, leaving five pRNAs still bound.[111] The pentameric theory is based on the fact that pRNA binds to the procapsid containing a pentameric vertex. However, recent publications have revealed that the foothold for the pRNA is the connector, not the procapsid protein.[149] Further scrutiny to argue against a stoichiometry of five involved in the DNA translocation process was accomplished by testing two sets of pRNAs. The first set was termed "Four Plus One" and the second was "Six Minus One." Both sets were found to be inactive in DNA packaging (manuscript in preparation).

"Four Plus One"

A key finding in pRNA research was that the mixing of two inactive pRNAs with complementary interlocking left and right loops restores DNA-packaging activity. For example, mixing two inactive pRNAs with *X-y′* and *Y-x′* resulted in production of infectious virions, which suggested that the right-hand loop *X* of pRNA *X-y′* can pair with the left-hand loop *x′* of pRNA *Y-x′*. Incorporation of one active monomer into the "Two Dimer Set" containing self-complementary loops can compose the "Four Plus One" set. For instance, mixing two pRNAs *A-a′* with monomer pRNA *Y-x′* can make up a five-member pRNA ring by trans-complementary interlocking. In this case, the Y and *x′* are the two loops that were complementary to the *a′* and *A*, respectively. Since monomer pRNA *Y-a′* shows significant DNA-packaging activity, right-hand loop *Y* can pair with left-hand loop *a′*. However, this combination for the formation of a five-pRNA ring (two plus two plus one) blocks the DNA packaging, indicating that the five-member pRNA ring formed by "Four Plus One" is inactive (Table 8).

Table 8. Engineered pRNA ring set to argue the pentameric pRNA ring theory, that are "Four Plus One" and "Six Minus One," respectively. The number inside ring represented yield of infectious virions (PFU/ml) in *in vitro* assembly assays (Shu, Zhang and Guo, manuscript in preparation)



"Six Minus One"

Another set, "Six Minus One," to compose a five-member pRNA ring, was accomplished by incorporating one inactive mutant dimer pRNA to replace one member of the two trans-complementary dimer pRNAs — that is, the "Three Dimer Set" (Table 8). For example, mixing two dimers with active (*A-b´*)(*B-a´*) and one mutant dimer

inactive (*A-b´*)(*B-a´*CCA) is expected to temporarily form a hexameric intermediate ring. As reported previously, mutant pRNA *B–a´*CCA, in which the $C_{18}C_{19}A_{20}$ bulge has been eliminated by insertion of three bases 5′-UGG between A_{99} and A_{100}, is competent in hexameric ring formation, but having no DNA packaging function.[150] If a five-member pRNA ring were derived from a hexameric intermediate, one pRNA should be dissociated from the intermediate hexameric ring to form a five-member ring, as suggested by Rossmann *et al.*[111] However, the "Six Minus One" set failed to produce infectious virions, which means that "Six Minus One" cannot produce active five-member pRNA rings.

As noted above, it is clearly revealed that five-member pRNA rings derived from either a "Four Plus One" or a "Six Minus One" approach were inactive in DNA packaging, arguing strongly against the hypothesis that five pRNA make up a pentamer to drive the phi29 DNA-packaging motor (manuscript in preparation). Thus it is not easy to account for how a pentameric pRNA ring could be formed from an even-numbered binding unit (pRNA dimer) at a targeted connector base that has even-numbered symmetry. This procedure could potentially be applicable to any system in which the possible stoichiometries need to be narrowed down and spurious options eliminated.

Perspectives

This review covers mathematical approaches that have been applied for stoichiometry quantification of the components of bacteriophage phi29 assembly system. All studies led to the same conclusion that phi29 DNA packaging motor contains six copies of pRNA, indicating that these methods are feasible. However, the reliability of these methods depends on the assay system. The more sensitive the assay system, the more accurate; the lower the background noise, the higher the reliability. For example, in phi29 the *in vitro* assembly system is highly sensitive, since in the presence of all components, up to 10^9 infectious virions can be assembled, and not a single plaque is detected if one of the components is omitted. The zero background

suggests that this system reaches nine order magnitude of sensitivity. When readers consider the application of these methods for their studies, the background noise level and the sensitivity of the system should be examined to assess the feasibility in their studies.

References

1. Anderson DL, Reilly B. (1993) Morphogenesis of bacteriophage ϕ29. In A.L. Sonenshein, J.A. Hoch, R. Losick (eds), *Bacillus Subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, pp. 859-867. American Society for Microbiology, Washington, D.C.
2. Bazinet C, King J. (1985) The DNA translocation vertex of dsDNA bacteriophages. *Ann Rev Microbiol* **39**: 109–129.
3. Black LW. (1989) DNA Packaging in dsDNA bacteriophages. *Ann Rev Microbiol* **43**: 267–292.
4. Casjens S, Hendrix R. (1988) Control mechanisms in dsDNA bacteriophage assembly. In R. Calendar (ed), *The Bacteriophages Vol.1*, pp. 15–92. Plenum Pubishing Corp., New York.
5. Guo P. (1994) Introduction: Principles, perspectives, and potential applications in viral assembly. *Seminars in Virology (Editor's Introduction)* **5(1)**: 1–3.
6. Cepko CL, Sharp PA. (1982) Assembly of adenovirus major capsid protein is mediated by a nonvirion protein. *Cell* **31**: 407–415.
7. D'Halluin JC, Milleville M, Boulanger PA, Martin GR. (1978) Temperature-sensitive mutants of adenovirus type 2 blocked in virion assembly: accumulation of light intermediate particles. *J Virol* **26**: 344–356.
8. Edvardsson B, Everitt E, Joernvall E, *et al.* (1976) Intermediates in adenovirus assembly. *J Virol* **19**: 533–547.
9. Gustin KE, Imperiale MJ. (1998) Encapsidation of viral DNA requires the adenovirus L1 52/55-kilodalton protein. *J Virol* **72**: 7860–7870.
10. Schmid SI, Hearing P. (1997) Bipartite structure and functional independence of adenovirus type 5 packaging elements. *J Virol* **71**: 3375–3384.
11. Abbotts AP, Preston VG, Hughes M, *et al.* (2000) Interaction of the herpes simplex virus type 1 packaging protein UL15 with full-length and deleted forms of the UL28 protein. *J Gen Virol* **81 Pt 12**: 2999–3009.

12. Dasgupta A, Wilson DW. (1999) ATP Depletion Blocks Herpes Simplex Virus DNA Packaging and Capsid Maturation. *J Virol.* **73(3)**: 2006–2015.
13. Rixon FJ, McNab D. (1999) Packaging-competent capsids of a herpes simplex virus temperature-sensitive mutant have properties similar to those of *in vitro*-assembled procapsids. *J Virol* **73**: 5714–5721.
14. DeLange AM, Reddy M, Scraba D, *et al.* (1986) Replication and resolution of cloned poxvirus telomeres *in vivo* generates linear minichromosomes with intact viral hairpin termini. *J Virol* **59**: 249–259.
15. Moss B. (1985) Replication of Poxviruses. In B.N. Fields, *et al.* (eds), *Virology*, pp. 685–703. Raven Press, New York.
16. Parsons BL, Pickup DJ. (1990) Transcription of orthopoxvirus telomeres at late times during infection. *Virology* **175**: 69–80.
17. Drexler H. (1984) Initiation by bacteriophage T1 of DNA packaging at a site between the P and Q genes of bacteriophage l. *J Virol* **49**: 754–759.
18. Shibata H, Fujisawa H, Minagawa T. (1987) Early events in a defined *in vitro* system for packaging of bacteriophage T3 DNA. *Virology* **159**: 250–258.
19. Black LW. (1981) *In vitro* packaging of bacteriophage T4 DNA. *Virology* **113**: 336–344.
20. Everett RD. (1981) DNA replication of bacteriophage T5. 3. Studies on the structure of concatemeric T5 DNA. *J Gen Virol* **52** (Pt 1): 25–38.
21. Cerritelli ME, Studier FW. (1996) Purification and characterization of T7 head-tail connectors expressed from the cloned gene. *J Mol Biol* **285**: 299–307.
22. Sun M, Louie D, Serwer P. (1999) Single-event analysis of the packaging of bacteriophage T7 DNA concatemers in vitro. *Biophys J* **77(3)**: 1627–1637.
23. Skorupski K, Sauer B, Sternberg N. (1994) Faithful cleavage of the P1 packaging site (pac) requires two phage proteins, PacA and PacB, and two Escherichia coli proteins, IHF and HU. *J Mol Biol* **243**: 268–282.
24. Pruss G, Calendar R. (1978) Maturation of bacteriophage P2 DNA. *Virology* **86**: 454–467.
25. Rishovd S, Holzenburg A, Johansen BV, Lindqvist BH. (1998) Bacteriophage P2 and P4 morphogenesis: structure and function of the connector. *Virology* **245**: 11–17.

26. Eppler K, Wyckoff E, Goates J, *et al.* (1991) Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging. *Virology* **183**: 519–538.
27. Greene B, King J. (1999) *In vitro* unfolding/refolding of wild type phage P22 scaffolding protein reveals capsid-binding domain. *J Biol Chem* **274**: 16135–16140.
28. Moore SD, Prevelige PE, Jr. (2001) Structural transformations accompanying the assembly of bacteriophage P22 portal protein rings *in vitro*. *J Biol Chem* **276** (9): 6779–6788.
29. Burns CM, Chan HLB, DuBow MS. (1990) *In vitro* maturation and encapsidation of the DNA of transposable μ-like phage D108. *Proc Natl Acad Sci USA* **87**: 6092–6096.
30. Smith MP, Feiss M. (1993) Sequence analysis of the phage 21 genes for prohead assembly and head completion. *Gene* **126**: 1–7.
31. Lee CS, Guo P. (1995) *In vitro* assembly of infectious virions of ds-DNA phage ϕ29 from cloned gene products and synthetic nucleic acids. *J Virol* **69**: 5018–5023.
32. Gutierrez C, Freire R, Salas M, Hermoso JM. (1994) Assembly of phage phi29 genome with viral protein p6 into a compact complex. *EMBO J* **13(1)**: 269–276.
33. Valpuesta JM, Fernandez JJ, Carazo JM, Carrascosa JL. (1999) The three-dimensional structure of a DNA translocating machine at 10 A resolution. *Structure Fold Des* **7**: 289–296.
34. Bjornsti MA, Reilly BE, Anderson DL. (1983) Morphogenesis of bacteriophage ϕ29 of *Bacillus subtilis*: oriented and quantized *in vitro* packaging of DNA protein gp3. *J Virol* **45**: 383–396.
35. Guo P, Grimes S, Anderson D. (1986) A defined system for *in vitro* packaging of DNA-gp3 of the *Bacillus subtilis* bacteriophage ϕ29. *Proc Natl Acad Sci USA* **83**: 3505–3509.
36. Gage LP, Geiduschek EP. (1971) RNA synthesis during bacteriophage SPO1 development: six classes of SPO1 RNA. *J Mol Biol* **57** (2): 279–297.
37. Levner MH, Cozzarelli NR. (1972) Replication of viral DNA in SPO1-infected Bacillus subtilis. I. Replicative intermediates. *Virology* **48** (2): 402–416.
38. Dubé P, Tavares P, Lurz R, van Heel M. (1993) The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J* **12**: 1303–1309.

39. Gual A, Camacho AG, Alonso JC. (2000) Functional analysis of the terminase large subunit, G2P, of Bacillus subtilis bacteriophage SPP1. *J Biol Chem* **275** (45): 35311–35319.
40. Becker A, Gold M. (1988) Prediction of an ATP reactive center in the small subunit, gpNul of phage lambda terminase enzyme. *J Mol Biol* **199**: 219–222.
41. Dokland T, Murialdo H. (1993) Structural transitions during maturation of bacteriophage lambda capsids. *J Mol Biol* **233**: 682–694.
42. Hang JQ, Catalano CE, Feiss M. (2001) The Functional Asymmetry of cosN, the Nicking Site for Bacteriophage lambda DNA Packaging, Is Dependent on the Terminase Binding Site, cosB. *Biochemistry* **40** (44): 13370–13377.
43. Woods L, Catalano CE. (1999) Kinetic characterization of the GTPase activity of phage lambda terminase: evidence for communication between the two "NTPase" catalytic sites of the enzyme. *Biochemistry* **Nov 2;38(44)**: 14624–14630.
44. Smith DE, Tans SJ, Smith SB, *et al.* (2001) The bacteriophage phi29 portal motor can package DNA against a large internal force. *Nature* **413**: 748–752.
45. Shu D, Guo P. (2003) Only one pRNA hexamer but multiple copies of the DNA-packaging protein gp16 are needed for the motor to package bacterial virus phi29 genomic DNA. *Virology* **309(1)**: 108–113.
46. Casjens S, Wyckoff E, Hayden M, *et al.* (1992) Bacteriophage P22 portal protein is part of the guage that regulates packing density of intravirion DNA. *J Mol Biol* **224**: 1055–1074.
47. Moore SD, Prevelige PE, Jr. (2002) DNA packaging: a new class of molecular motors. *Curr Biol* **12** (3): R96–8.
48. Waterman MS, Gordon L, Arratia R. (1987) Phase transitions in sequence matches and nucleic acid structure. *Proc Natl Acad Sci USA* **84**: 1239–1243.
49. White JH, Cozzarelli NR, Bauer WR. (1988) Helical repeat and linking number of surface-wrapped DNA. *Science* **241**: 323–327.
50. Cohen FE, Abarbanel RA, Kuntz ID, Fletterick RJ. (1986) Turn prediction in proteins using a pattern-matching approach. *Biochemistry* **25**: 266–275.
51. Becker A, Marko M, Gold M. (1977) Early events in the *in vitro* packaging of Bacteriophage DNA. *Virology* **78**: 291–305.

52. Israel JV, Anderson TH, Levine M. (1967) *In vitro* morphogenesis of phage p22 from heads and base-plate parts. *Proc Natl Acad Sci USA* **57**: 284–291.
53. Prevelige PEJ, Thomas D, King J. (1988) Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells *in vitro*. *J Mol Biol* **202**: 743–757.
54. Thomas D, Prevelige P. (1991) A Pilot Protein Participates in the Initiation of P22 Procapsid Assembly. *Virology* **182**: 637–681.
55. Earnshaw WC, Casjens SR. (1980) DNA packaging by the double-stranded DNA bacteriophages. *Cell* **21**: 319–331.
56. Niemeyer CM. (2002) The developments of semisynthetic DNA-protein conjugates. *Trends Biotechnol* **20** (9): 395–401.
57. Schmidt OG, Eberl K. (2001) Nanotechnology. Thin solid films roll up into nanotubes. *Nature* **410**: 168.
58. Baneyx G, Baugh L, Vogel V. (2002) Supramolecular Chemistry and Self-assembly Special Feature: Fibronectin extension and unfolding within cell matrix fibrils controlled by cytoskeletal tension. *Proc Natl Acad Sci USA* **99** (8): 5139–5143.
59. Hyman P, Valluzzi R, Goldberg E. (2002) Design of protein struts for self-assembling nanoconstructs. *Proc Natl Acad Sci USA* **99** (13): 8488–8493.
60. Goldberger J, He R, Zhang Y, *et al.* (2003) Single-crystal gallium nitride nanotubes. *Nature* **422**: 599–602.
61. D'Halluin J-C, Martin GR, Torpier G, Boulanger PA. (1978) Adenovirus type 2 assembly analyzed by reversible crosslinking of labile intermediates. *J Virol* **26**: 357–363.
62. Zandonella C. (2003) Cell nanotechnology: The tiny toolkit. *Nature* **423**: 10–12.
63. Yildiz A, Forkey JN, McKinney SA, *et al.* (2203) Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* **300**: 2061–2065.
64. Serwer P. (2003) Models of bacteriophage DNA packaging motors. *J Struct Biol* **141** (3): 179–188.
65. Oster G, Wang H. (2003) Energy transduction in the F1 motor of ATP synthase. *Nature* **396(6708)**: 279–282.
66. Berry RM. (2003) Theories of rotary motors. *Philos Trans R Soc Lond B Biol* **355(1396)**: 503–509.

67. Grigoriev DN, Moll W, Hall J, Guo P. (2003) Bionanomotor. *Encyclopedia of Nanoscience and Nanotechnology* **1**: 361–374.
68. Inoue A, Saito J, Ikebe R, Ikebe M. (2002) Myosin IXb is a single-headed minus-end-directed processive motor. *Nat Cell Biol* **4** (4): 302–306.
69. Sablin EP, Fletterick RJ. (2001) Nucleotide switches in molecular motors: structural analysis of kinesins and myosins. *Curr Opin Struct Biol* **11** (6): 716–724.
70. Guo Y, Blocker F, Xiao F, Guo P. (2005) Construction and 3-D computer modeling of connector arrays with tetragonal to decagonal transition induced by pRNA of phi29 DNA-packaging motor. *Journal of Nanosci and Nanotech (JNN)* **5**: 856–863.
71. Shenton W, Pum D, Sleytr UB, Mann S. (1997) Synthesis of cadmium sulphide superlattices using self-assembled bacterial S-layers. *Nature* **389**: 585–587.
72. Carazo J, Santisteban A, Carrascosa J. (1985) Three-dimensional Reconstruction of Bacteriophage ϕ29 Neck Particles at 2.2 nm Resolution. *J Mol Biol* **183**: 79–88.
73. Jimenez J, Santisteban A, Carazo JM, Carrascosa JL. (1986) Computer graphic display method for visualizing three-dimensional biological structures. *Science* **232**: 1113–1115.
74. Dujardin E, Peet C, Stubbs G, *et al.* (2003) Organization of Metallic Nanoparticles Using Tobacco Mosaic Virus Templates. *Nano Letters* **3 (3)**: 413–417.
75. Soong RK, Bachand GD, Neves HP, *et al.* (2000) Powering an inorganic nanodevice with a biomolecular motor. *Science* **290** (5496): 1555–1558.
76. Hoeprich S, ZHou Q, Guo S, *et al.* (2003) Bacterial virus phi29 pRNA as a hammerhead ribozyme escort to destroy hepatitis B virus. *Gene Therapy* **10(15)**: 1258–1267.
77. Hess H, Vogel V. (2001) Molecular shuttles based on motor proteins: Active tranport in synthetic environments. *Reviews in Mol Biotechn* **82** (1): 67–85.
78. Craighead HG. (2000) Nanoelectromechanical systems. *Science* **290**: 1532–1536.
79. Fennimore AM, Yuzvinsky TD, Han WQ, *et al.* (2003) Rotational actuators based on carbon nanotubes. *Nature* **424**: 408–410.

80. Wall JS, Hainfeld JF, Simon MN. (1998) Scanning transmission electron microscopy of nuclear structures. *Methods Cell Biol* **53**: 139–164.
81. Nilsson J, Miyazaki N, Xing L, *et al.* (2005) Structure and assembly of a T=1 virus-like particle in BK polyomavirus. *J Virol* **79**: 5337–5345.
82. Frank J. (2001) Cryo-electron microscopy as an investigative tool: the ribosome as an example. *BioEssays* **23**: 725–732.
83. Baker TS, Olson NH, Fuller SD. (1999) Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol Mol Biol Rev* **63**: 862–922, table.
84. Balzani V, Credi A, Venturi M. (2002) The bottom-up approach to molecular-level devices and machines. *Chemistry* **8** (24): 5524–5532.
85. Seeman NC, Belcher AM. (2002) Emulating biology: building nanostructures from the bottom up. *Proc Natl Acad Sci USA* **99 Suppl 2**: 6451–6455.
86. Murialdo H, Becker A. (1977) Assembly of biologically active proheads of bacteriophage lambda *in vitro*. *Proc Natl Acad Sci USA* **74**: 906–910.
87. Kerr C, Sadowski PD. (1974) Packaging and maturation of DNA of bacteriophage T7 *in vitro*. *Proc Natl Acad Sci USA* **71**: 3545–3549.
88. Pruss GJ, Wang JC, Calendar R. (1975) *In vitro* packaging of covalently closed monomers of bacteriophage DNA. *J Mol Biol* **98**: 465–478.
89. Kaiser AD, Syvanen M, Masuda T. (1975) DNA packaging steps in bacteriophage lambda head assembly. *J Mol Biol* **91**: 175–186.
90. Bancroft JB, Hills GJ, Markham R. (1967) A study of the self-assembly process in a small spherical virus. Formation of organized structures from protein subunits *in vitro*. *Virology* **31**: 354.
91. King J, Casjens S. (1974) Catalytic head assembly protein in virus morphogenesis. *Nature* **251**: 112–119.
92. Leibowitz J, Horwitz MS. (1975) Synthesis and assembly of adenovirus polypeptieds. III. Reversible inhibition of hexon assembly at 42 C. *Virology* **66**: 10–24.
93. Hsiao CL, Black LW. (1977) DNA packaging and the pathway of bacteriophage T4 head assembly. *Proc Natl Acad Sci USA* **74**: 3652–3656.
94. Skehel J, Joklik W. (1969) Studies on the *in vitro* transcription of reovirus RNA catalyzed by reovirus cores. *Virology* **39**: 822–831.

95. Wilhelm J, Ginsberg HS. (1972) Synthesis *in vitro* of type 5 adenovirus capsid proteins. *J Virol* **9**: 973–980.
96. Kaiser D, Masuda T. (1973) *In vitro* assembly of bacteriophage lambda heads. *Proc Natl Acad Sci USA* **70**: 260–264.
97. Hohn B. (1975) DNA as substrate for packaging into bacteriophage lambda, *in vitro*. *J Mol Biol* **98**: 93–106.
98. Miyazaki J, Fugisawa H, Minagawa T. (1978) Biological activity of purified bacteriophage T3 prohead and prohead-like stuctures as precusors for *in vitro* head assembly. *Virology* **91**: 283–290.
99. Hill AV. (1910) The combinations of haemoglobin with oxygen and with carbon monoxide. *J Physiol* **40**: iv–vii.
100. Yifrach O. (2004) Hill coefficient for estimating the magnitude of cooperativity in gating transitions of voltage-dependent ion channels. *Biophys J* **87**: 822–830.
101. Stanca SE, Popescu IC. (2004) Phenols monitoring and Hill coefficient evaluation using tyrosinase-based amperometric biosensors. *Bioelectrochemistry* **64**: 47–52.
102. Hofer GF, Hohenthanner K, Baumgartner W, *et al*. (1997) Intracellular Ca2+ inactivates L-type Ca2+ channels with a Hill coefficient of approximately 1 and an inhibition constant of approximately 4 microM by reducing channel's open probability. *Biophys J* **73**: 1857–1865.
103. Edelstein SJ, Bardsley WG. (1997) Contributions of individual molecular species to the Hill coefficient for ligand binding by an oligomeric protein. *J Mol Biol* **267**: 10–16.
104. Monod J, Wyman J, Changeux J. (1965) On the nature of allosteric transitions: a plausible model. *J Mol Biol* **12**: 88–118.
105. Weiss JN. (1997) The Hill equation revisited: uses and misuses. *The FASEB Journal* **11**: 835–841.
106. Chen C, Sheng S, Shao Z, Guo P. (2000) A Dimer as a Building Block in Assembling RNA. A hexamer that gears bacterial virus phi29 DNA-translocating machinery. *J Biol Chem* **275(23)**: 17510–17516.
107. Trottier M, Guo P. (1997) Approaches to determine stoichiometry of viral assembly components. *J Virol* **71**: 487–494.
108. Chen C, Trottier M, Guo P. (1997) New approaches to stoichiometry determination and mechanism investigation on RNA involved in intermediate reactions. *Nucleic Acids Symposium Series* **36**: 190–193.

109. Guasch A, Pous J, Ibarra B, *et al.* (2002) Detailed architecture of a DNA translocating machine: the high- resolution structure of the bacteriophage phi29 connector particle. *J Mol Biol* **315** (4): 663–676.
110. Simpson AA, Leiman PG, Tao Y, *et al.* (2001) Structure determination of the head-tail connector of bacteriophage phi29. *Acta Cryst* **D57**: 1260–1269.
111. Simpson AA, Tao Y, Leiman PG, *et al.* (2000) Structure of the bacteriophage phi29 DNA packaging motor. *Nature* **408** (6813): 745–750.
112. Ibarra B, Caston JR, Llorca O, *et al.* (2000) Topology of the components of the DNA packaging machinery in the phage phi29 prohead. *J Mol Biol* **298** (5): 807–815.
113. Guo P, Peterson C, Anderson D. (1987) Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage ϕ29. *J Mol Biol* **197**: 229–236.
114. Hagen EW, Reilly BE, Tosi ME, Anderson DL. (1976) Analysis of gene function of bacteriophage ϕ29 of *Bacillus subtilis*: identification of cistrons essential for viral assembly. *J Virol* **19(2)**: 501–517.
115. Guo P, Erickson S, Anderson D. (1987) A small viral RNA is required for *in vitro* packaging of bacteriophage ϕ29 DNA. *Science* **236**: 690–694.
116. Shu D, Guo P. (2003) A Viral RNA that binds ATP and contains an motif similar to an ATP-binding aptamer from SELEX. *J Biol Chem* **278(9)**: 7119–7125.
117. Huang LP, Guo P. (2003) Use of acetone to attain highly active and soluble DNA packaging protein gp16 of phi29 for ATPase assay. *Virology* **312(2)**: 449–457.
118. Grimes S, Anderson D. (1990) RNA Dependence of the Bateriophage phi29 DNA Packaging ATPase. *Mol Biol* **215**: 559–566.
119. Guo P, Bailey S, Bodley JW, Anderson D. (1987) Characterization of the small RNA of the bacteriophage ϕ29 DNA packaging machine. *Nucleic Acids Res* **15**: 7081–7090.
120. Bailey S, Wichitwechkarn J, Johnson D, *et al.* (1990) Phylogenetic analysis and secondary structure of the *Bacillus subtilis* bacteriophage RNA required for DNA packaging. *J Biol Chem* **265**: 22365–22370.
121. Chen C, Zhang C, Guo P. (1999) Sequence requirement for hand-in-hand interaction in formation of pRNA dimers and hexamers to gear phi29 DNA translocation motor. *RNA* **5**: 805–818.

122. Chen C, Guo P. (1997) Magnesium-induced conformational change of packaging RNA for procapsid recognition and binding during phage phi29 DNA encapsidation. *J Virol* **71**: 495–500.
123. Reid RJD, Bodley JW, Anderson D. (1994) Characterization of the prohead-pRNA interaction of bacteriophage ϕ29. *J Biol Chem* **269**: 5157–5162.
124. Chen C, Guo P. (1997) Sequential action of six virus-encoded DNA-packaging RNAs during phage phi29 genomic DNA translocation. *J Virol* **71**: 3864–3871.
125. Chen C, Guo P. (1997) Sequential action of six virus-encoded DNA-packaging RNAs during phage phi29 genomic DNA translocation. *J Virol* **71** (5): 3864–3871.
126. Garver K, Guo P. (1997) Boundary of pRNA functional domains and minimum pRNA sequence requirement for specific connector binding and DNA packaging of phage phi29. *RNA* **3**: 1068–1079.
127. Reid RJD, Zhang F, Benson S, Anderson D. (1994) Probing the structure of bacteriophage ϕ29 prohead RNA with specific mutations. *J Biol Chem* **269**: 18656–18661.
128. Reid RJD, Bodley JW, Anderson D. (1994) Identification of bacteriophage ϕ29 prohead RNA (pRNA) domains necessary for *in vitro* DNA-gp3 packaging. *J Biol Chem* **269**: 9084–9089.
129. Zhang CL, Lee C-S, Guo P. (1994) The proximate 5' and 3' ends of the 120-base viral RNA (pRNA) are crucial for the packaging of bacteriophage ϕ29 DNA. *Virology* **201**: 77–85.
130. Zhang CL, Tellinghuisen T, Guo P. (1995) Confirmation of the helical structure of the 5'/3' termini of the essential DNA packaging pRNA of phage ϕ29. *RNA* **1**: 1041–1050.
131. Trottier M, Zhang CL, Guo P. (1996) Complete inhibition of virion assembly *in vivo* with mutant pRNA essential for phage ϕ29 DNA packaging. *J Virol* **70**: 55–61.
132. Lee CS, Guo P. (1994) A highly sensitive system for the *in vitro* assembly of bacteriophage ϕ29 of *Bacillus subtilis. Virology* **202**: 1039–1042.
133. Guo P, Zhang C, Chen C, *et al.* (1998) Inter-RNA interaction of phage phi29 pRNA to form a hexameric complex for viral DNA transportation. *Mol Cell* **2**: 149–155.
134. Guo P. (2002) Methods for Structural and Functional Analysis of an RNA Hexamer of Bacterial Virus phi29 DNA Packaging motor. *ACTA BIOCHIMICA et BIOPHYSICA SINCA* **34(5)**: 533–543.

135. Trottier M, Mat-Arip Y, Zhang C, *et al.* (2000) Probing the structure of monomers and dimers of the bacterial virus phi29 hexamer RNA complex by chemical modification. *RNA* **6**: 1257–1266.
136. Garver K, Guo P. (2000) Mapping the inter-RNA interaction of phage phi29 by site-specific photoaffinity crosslinking. *J Biol Chem* **275(4)**: 2817–2824.
137. Zhang F, Lemieux S, Wu X, *et al.* (1998) Function of hexameric RNA in packaging of bacteriophage phi29 DNA in vitro. *Mol Cell* **2**: 141–147.
138. Peterson C, Simon M, Hodges J, *et al.* (2001) Composition and Mass of the Bacteriophage phi29 Prohead and Virion. *J Struct Biol* **135**: 18–25.
139. Tosi ME, Reilly BE, Anderson DL. (1975) Morphogenesis of bacteriophage ϕ29 of *Bacillus subtilis*: cleavage and assembly of the neck appendage protein. *J Virol* **16**: 1282–1295.
140. Villanueva N, Lazaro JM, Salas M. (1981) Purification, properties and assembly of the neck-appendage protein of the *Bacillus subtilis* phage phi 29. *Eur J Biochem* **117**: 499–505.
141. Villanueva N, Salas M. (1981) Adsorption of Bacteriophage phi 29 to *Bacillus subtilis* through the neck appendage of the viral particle. *J Virol* **38(1)**: 15–19.
142. Hendrix RW. (1998) Bacteriophage DNA packaging: RNA gears in a DNA transport machine (Minireview). *Cell* **94**: 147–150.
143. Tao Y, Olson NH, Xu W, *et al.* (1998) Assembly of a Tailed Bacterial Virus and Its Genome Release Studied in Three Dimensions. *Cell* **95**: 431–437.
144. Nancy Bourassa, Francois Major. (2002) Implication of the prohead RNA in phage phi29 DNA packaging. *Biochemie* **84**: 945–951.
145. Ibarra B, Caston JR, Llorca O, *et al.* (2000) Topology of the components of the DNA packaging machinery in the phage phi29 prohead. *J Mol Biol* **298**: 807–815.
146. Hoeprich S, Guo P. (2002) Computer Modeling of Three-dimensional Structure of DNA-packaging RNA(pRNA) Monomer, Dimer, and Hexamer of Phi29 DNA Packaging Motor. *J Biol Chem* **277(23)**: 20794–20803.
147. Morais MC, Choi KH, Koti JS, *et al.* (2005) Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of phi29. *Mol Cell* **18**: 149–159.

148. Morais MC, Tao Y, Olsen NH, *et al.* (2001) Cryoelectron-Microscopy Image Reconstruction of Symmetry Mismatches in Bacteriophage phi29. *J Struct Biol* **135**: 38–46.
149. Xiao F, Moll D, Guo S, Guo P. (2005) Binding of pRNA to the N-terminal 1amino acids of connector proteinof bacterial phage phi29. *Nucleic Acids Res* **33, No.8**: 2640–2649.
150. Zhang CL, Tellinghuisen T, Guo P. (1997) Use of circular permutation to assess six bulges and four loops of DNA-Packaging pRNA of bacteriophage ϕ29. *RNA* **3**: 315–322.
151. Shu D, Huang L, Guo P. (2004) A simple mathematical formula for stoichiometry quantification of viral and nanobiological assemblage using slopes of log/log plot curves. *J Virol Meth* **115**: 19–30.

Chapter 16

Virus-like Particles of Fish Nodavirus

*Chan-Shing Lin**

Betanodaviruses are causative agents of neurological disorders in several species of fish. I reviewed the existing research on the structure of the fish nodavirus. We sequenced and expressed the RNA2 segment of two grouper viruses isolated from *Epinephelus malabaricus* (malabaricus grouper nervous necrosis virus, MGNNV) and *Epinephelus lanceolatus* (dragon grouper nervous necrosis virus, DGNNV). Comparing the ORF with striped jack nervous necrosis virus, Atlantic halibut virus, sea bass encephalitis virus, and greasy grouper nervous necrosis virus (GGNNV), the analysis revealed that the Japanese SJNNV differed significantly from its geographical neighbors, the Taiwanese MGNNV and DGNNV, particularly in the variable region. In contrast, the geographically proximal European DlEV and AHV showed a higher degree of similarity with each other and the Taiwanese viruses. This suggests that host-dependent selection impacts more significantly on evolution than geographically related conditions such as water temperature. The capsid proteins of MGNNV and DGNNV can form virus-like particles (VLPs) in Sf21 cells with a recombinant baculovirus system and *E. coli*. An image reconstruction with electron cryomicroscopy data revealed a morphology of the VLPs consistent with $T = 3$ quasi-symmetry, but with features significantly different from insect nodavirus structures at the same resolution. This assembly system allows the first biophysical comparisons of fish and insect nodavirus structure, assembly, and stability. The cryoEM structure at 23 Å resolution, sequence comparison, and protein fold recognition

* The Department of Marine Biotechnology and Resources, National Sun Yat-sen University, Kaohsiung, TAIWAN 804.

analysis indicate that the coat protein of MGNNV has two domains resembling those of tomato bushy stunt virus and Norwalk virus, rather than the expected single-domain coat protein of insect nodaviruses. The analysis implies that residues 83 to 216 fold as a β-sandwich that forms the inner shell of the $T = 3$ capsid and residues 217 to 308 form the trimeric surface protrusions observed in the cryoEM map. Deleting N- and C-termini revealed different impacts on VLP formation. Deletion of 35 or 52 residues at the N-terminus completely ruined the VLP assembly, presumably due to removal of positively charged residues for binding RNAs. When deletions were restricted to 4, 16, or 25 N-terminal residues, the assembly of VLPs remained. The ability of VLP formation diminished when 4 to 11 C-terminal residues were deleted. The termini that can be deleted without seriously destructing the VLPs are 25 and 3 residues at N- and C-termini, respectively. The recombinant VLPs block attachment of native virus to the surface of cultured fish cells, thereby blocking infection by the native virus.

Introduction

After five years of observations in the Kyushu hatchery, Mori *et al.* published in 1992 the new virus, a spherical particle from infected larval striped jack (*Pseudocaranx dentex*), and classified it as a nodavirus by its bipartite geneme of ~1.5 kb and ~3 kb.[1] Using cell-free extract of rabbit reticulocytes, three bands of proteins were found, which were 40 kDa, 42 kDa, and 100 kDa. The protein of 42 kDa is capsid, while RNA polymerase is 100 kDa; 40 kDa was presumed as a truncated capsid protein. The virus particles were mainly found in brain and retinal cells, with a diameter of 25–30 nm and with non-envelope icosahedral shape. This is the first *Betanovirus* identification that was once mistaken as picornavirus.

Piscine nodaviruses were officially classified in the genus *Betanodavirus* by ICV in 2001, although Asian scientists once rather named it as piscinodavirus. All of the literature mentioned in this paper can be referenced to the review by Munday *et al.*[2] Nodaviridae includes two genuses: insect *Alphanodavirus* and fish *Betanodavirus*. The viral disease is found in many species of marine fish in Asia, Europe and Australia. The symptoms of the diseased fish are described as viral encephalopathy and retinopathy (VER), or viral nervous necrosis (VNN), accompanied by syndromes of abnormal swimming behavior, darkness of skin, and massive mortality in hatchery-reared larvae and

juveniles. Once again, Asian scientists tend to use VNN to describe the disease symptom, while Western scientists tend to use VER. The concept differences between names may be clarified by whether the nerve cells are affected, which has yet to be confirmed by any firm evidence.

Genome of Fish Nodavirus

The genome of the nodaviruses is bipartite of single-stranded, positive-sense RNAs: RNA1 assumedly encodes the viral replicase, whereas RNA2 encodes the capsid protein. The complete sequence of RNA1 and RNA2 of striped jack nervous necrosis virus (SJNNV) was determined in 1995 and 1999, respectively, while we completed the RNA2 of grouper NNV in 1999 and RNA1 in 2001. The sequences of RNA2 of *Dicentrarchus labrax* encephalitis virus (DlEV) and Atlantic halibut virus (AHV) (Grotmol *et al.,* 2000) were also determined.

RNA2 of MGNNV and DGNNV, two grouper nervous necrosis viruses isolated from Taiwanese *Epinephelus malabaricus* and *Epinephelus lanceolatus,* respectively, was cloned and sequenced using standard procedures (Lu and Lin, 1999).[3] Both RNAs were 1390 nucleotides (nt) long and contained a single open reading frame (ORF) encoding a 338-amino-acid protein of approximately 37 kDa, the putative capsid protein. The ORF spanned nt 15–1031, leaving an untranslated region of 14 nt at the 5′-end and 359 nt at the 3′-end. Alignment of the two sequences showed that they were 99% identical at both the nucleotide level and the amino acid level. Comparison with the recently determined sequence of RNA2 from the grouper virus GGNNV indicated that the RNA2 of MGNNV and DGNNV contained shorter 5′ and 3′ non-coding regions. This may be due to the absence of authentic 5′ and 3′ termini in the MGNNV and DGNNV clones, which were generated with primers based on the 5′- and 3′-end sequences of SJNNV RNA2. Despite the discrepancy in length, RNA2 of all three grouper viruses contained an ORF encoding a protein of 338 amino acids.

Based upon T4-region sequences from more than 20 other betanodavirus RNA2 segments, Nishizawa *et al.* (1997) classified the betanodaviruses into four groups: SJNNV, tiger puffer NNV (TPNNV), barfin flounder NNV (BFNNV), and red grouper NNV (RGNNV).

Comparative analysis of the coat protein genes of five betanodaviruses has shown that the sequence can be divided into a highly conserved region and a variable region. In SJNNV, the conserved region is represented by amino acids 83–216 and the variable region by amino acids 235–315. Based on this division, the sequences of MGNNV and DGNNV were compared at the nucleotide level and amino acid level with the sequences of four other betanodaviruses of Asian and European origin.[3] These viruses were SJNNV from Japan, DlEV from France, AHV from Norway, and GGNNV from Singapore. MGNNV and DGNNV viruses were most closely related to GGNNV. Interestingly, the analysis also revealed that the Japanese SJNNV differed significantly from its geographical neighbors, the Taiwanese MGNNV and DGNNV, particularly in the variable region. In contrast, the geographically proximal European DlEV and AHV showed a higher degree of similarity with each other and the Taiwanese viruses. This suggests that host-dependent selection impacts more significantly on evolution than geographically related conditions such as water temperature.[3]

MGNNV and DGNNV have a high concentration of basic amino acids in the N-terminal 50 residues. Insect nodaviruses contain 15–18 positively charged amino acids, mostly arginines, in the first 50 amino acids, while MGNNV and DGNNV contain 9 arginine and 6 lysine residues.[3] It is likely that the N-terminal basic arm of the fish virus coat protein is involved in binding and neutralization of the packaged RNA and that it is located inside the viral capsid as observed for the insect viruses. The arginine-rich region that can bind nucleotides with ionic bonds was coincidently considered as a nuclear internalization signal, which will be arbitrary possible.

Virus-like Particle Formation

The ORF region of RNA2 encoding the capsid protein of MGNNV was subcloned into a transfer vector and recombinant baculoviruses were generated using standard protocols.[3] Sf21 cells were then infected with the recombinant virus and lysed three days post-infection, and putative particles in the lysate were concentrated by pelleting through

a 20% sucrose cushion. SDS-polyacrylamide gel electrophoresis showed that the re-suspended pellet contained large amounts of a 37 kDa capsid protein that was not observed in mock-infected cells, presumably VLPs. The re-suspended pellet was centrifuged through a 10–40% (wt/wt) sucrose gradient followed by fractionation of the gradient with continuous absorbance at 254 nm. The absorbance profile showed three peaks, but gel electrophoresis of the corresponding fractions indicated that only the material in the two fastest peaks, labeled I and II, contained protein. This protein migrated as a single band and had an apparent molecular mass of approximately 37 kDa. Electron microscopy of negatively stained material taken from peak I showed the presence of irregularly assembled particles, many of which were stain-permeable, whereas peak II showed intact, non-permeable particles that had a diameter of approximately 31 nm. A comparison of these particles with authentic MGNNV virions confirmed that the VLPs were closely similar in size and geometry to native virus. Interestingly, the surface of both native particles and VLPs was distinctly rougher than that of the insect nodavirus Flock house virus (FHV). There was no evidence for the presence of coat protein cleavage products as observed for the insect nodaviruses. Neither of such cleavage products was detected by employing the more sensitive technique of mass spectrometry using purified VLPs (Lin, unpublished data). It was shown for DlEV that the protein doublet is not derived from an assembly-dependent maturation cleavage of the capsid coat protein as it is in the insect nodaviruses. Rather, it was proposed that the doublet might arise from cotranslational modification.

Piscine nodavirus VLPs package random cellular RNAs when assembled in Sf21 cells.[3] The particles in peak I contained a mixture of RNAs ranging in size from approximately 100 bases, presumably tRNAs, to 1300 bases, whereas particles in peak II contained a broader mixture of RNAs ranging in size from about 100 to 4500 bases. The observation was essentially identical to what has been seen for the insect nodavirus VLPs. The sharp cutoff in size probably reflects the largest single RNA molecule that can be packaged in this size of particle. The promiscuity of RNA packaging is likely to reflect the electrostatic requirements of assembly and the ability of the capsid protein to

recruit a variety of cellular RNA molecules to satisfy the highly charged amino-terminal portions of the subunit. The peak I particles were irregular because the coat protein had recruited RNA molecules too small to satisfy the requirements for forming an icosahedral lattice of the correct size.

The RNA2 of DGNNV includes a single ORF that encodes a 338 amino acid protein of 37,085 Da, the putative capsid protein. The recombinant clone is the first attempt to express nodavirus VLPs in prokaryote *E. coli*. The molecular weight of the recombinant full-length capsid protein on the VLPs was the same as the native virus coat protein. Electron microscopy revealed that the features of VLPs resemble the native virus. VLP formation from various viral capsid proteins expressed in *E. coli* has been reported in other types of viruses elsewhere. Two visible bands appeared at a density of $1.27\,g\,cm^{-3}$ (Fraction I) and $1.34\,g\,cm^{-3}$ (Fraction II). The concentration of Fraction I was approximately 5% of Fraction II, estimated by UV absorbance (A280). The molecular mass of the proteins (~37 kDa) in both fractions was apparently the same as the native virus in SDS-PAGE. Heterogeneous VLPs existed in both fractions. Electron microscopy of VLPs in Fraction I showed that the small particles were about 23 nm in diameter, a few of which were irregular. In contrast, Fraction II consisted of equal amounts of intact and stain-permeable particles with a diameter of about 30 nm. Comparing these particles revealed that the heavy VLPs had a similar size and geometry to the native piscine nodavirus and Sf21-expressed VLPs. Interestingly, the surfaces of both native particles and VLPs are rougher than that of the insect nodavirus flock house virus.[3] Similarly, two types of *Dicentrarchus labrax* encephalitis virus have been reported.

VLPs from N-terminus Deletion Mutants

The coat protein of DGNNV consists of a high percentage of basic residues in the N-terminal region, nine Arg and six Lys (Lu and Lin, 1999; AF245004, GenBank). The N-terminal basic arm is involved in binding the negatively charged phosphate backbone of the RNAs located inside the viral capsid. Several N-terminal deletion mutants of

the DGNNV capsid protein were first constructed to examine the involvement of the N-terminus in assembly of the virus.[4] The mutant ΔN52 (truncating 1–52 N-terminal residues) deleted all positive charges at the N-terminus. The mutant ΔN35 (truncating 1–35 N-terminal residues) deleted 13 positively charged residues out of 15, while the mutant ΔN25 (truncating 1–25 N-terminal residues) removed nine Lys/Arg. The main difference between ΔN25 and ΔN35 is the number of triple-Arg: ΔN25 retains one triple-Arg while ΔN35 has none. The ΔN52 and ΔN35 did not produce detectable VLPs while ΔN25 maintained the formation of VLPs, suggesting that the residues between the 25th and 35th residues are essential for forming VLPs. Among the 10 residues, the role of triple-Arg at the 29–31 residues in forming VLPs is noticeable. The density of ΔN25 VLPs is 1.34 g/cm^3, which is the same as the native particles. Further deletions within 25 N-terminal residues were performed to examine their functions in VLP formation. There is a triple-Arg located at 23–25 residues. With triple-Arg at both 23–25 and 29–31 residues, the mutants of deleting 4 (ΔN4) and 16 (ΔN16) N-terminal residues retained the formation of VLPs. These VLPs were very similar in shape to the VLPs of the full-length coat protein, but they were slightly smaller. The particle diameters of ΔN4 and ΔN25 were 26 and 24 nm, respectively. The quantities of VLPs from the N-terminal deletion mutants of ΔN4, ΔN16, and ΔN 25 are the same as that of the full-length capsid protein.

Expression of C-terminus Deletion Mutants

A mutant of deleting 289–338 C-terminal residues (ΔC288) was initially constructed where the site of Asn/Ala resembles the autocleavage sequence in alphanodavirus. The ΔC288 mutant produced a 31.5 kDa protein, but no VLPs were found. The mutants of deleting 328–338 (ΔC327) and 336–338 (ΔC335) C-terminal residues produced slightly smaller particles than those of native virus particles, while ΔC292, ΔC296, and ΔC308 did not produce VLPs.[4] The particle diameters of ΔC327 and ΔC335 were 24 nm and 26 nm, respectively. The particles with short truncations (ΔN4 and ΔC335)

are bigger than ΔN25 and ΔC327. The quantity of ΔC335 VLPs was 80% of full-length VLPs, while the amount of ΔC327 VLPs sharply decreased to 4.5%. The result indicates that the 328–335 residues play a crucial role in assembling VLPs. The relative percentages of VLPs from ΔC327, ΔC328, ΔC330, ΔC333, ΔC334, and ΔC335 indicate that the ability to form VLPs declined as successive residues were removed. Particularly, deleting four residues at 335–338 (ΔC334) significantly reduced the quantity of VLPs to 39.4%, while removing three residues at 336–338 (ΔC335) only reduced the percentage to 80%. An excess of deleting one residue at 335th (Asp-335) decreases VLPs to 39.4% from 80% of full-length capsid protein.

Electron Cryomicroscopy and Image Reconstruction

In order to obtain higher-resolution structural information about the fish virus VLPs, particles were flash-frozen for analysis by low-dose electron cryomicroscopy.[5] A total of 263 particles were used to generate a three-dimensional reconstruction at 23 Å resolution (Fig. 1). The particle had a diameter of 380 Å and the morphology was consistent with the expected $T = 3$ surface lattice previously observed for insect nodaviruses. As in the insect viruses, prominent protrusions were clearly visible at the quasi-threefold symmetry axes, but overall the surface features were strikingly different. For example, the protrusion at the quasi-threefold axis was significantly larger than that found in the insect viruses Flock house virus and Pariacoto virus

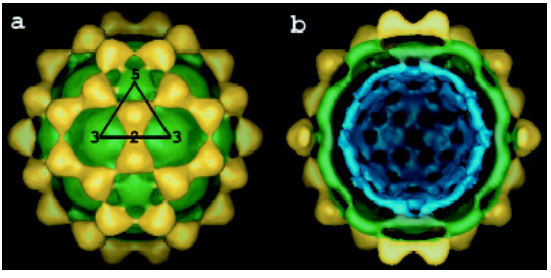


Fig. 1. CryoEM maps of MGNNV at 23 Å resolution.

(Reprinted from Tang *et al.*, 2002.)

and requires a larger insertion of polypeptides between strands of the β-sandwich if the subunit has the same overall structure as that of the insect nodaviruses. In fact, the striking gaps between the inner and the outer regions of the capsid are more consistent with a two-domain subunit than the integrated insertions seen in insect nodaviruses. The organization of conserved (residues 83–216) and variable (residues 235–315) regions of the subunit leads to a similar conclusion when it is compared to coat protein sequence alignments of insect nodaviruses where there are no such linear regions of conservation and variability. It was shown that the surface portion of the insect nodavirus subunits is considerably more variable in sequence than the β-sandwich, but since this region is formed by a variety of insertions between strands of the β-sheets, it does not stand out in the sequence alignments. In contrast, sequence comparison of the insect tetraviruses shows a striking region of variability that is linear and corresponds to the insertion of an entire domain between two strands of the β-sandwich that forms the $T = 4$ particle shell. Thus the linearity of the variable domain in the fish nodavirus coat protein suggests that it may correspond to a contiguous external domain in fish virus particles, thus indicating a different subunit organization than observed in the insect nodaviruses. Such an organization would be more similar to that of the coat protein subunits of the plant virus tomato bushy stunt virus, where residues 104 to 268 are in the contiguous β-sandwich shell-forming domain while residues 273–386 form the protruding domain on the particle outer surface.

The details of the morphology and the radial density distribution of MGNNV differ dramatically from those of PaV. The maximum diameter of MGNNV is approximately 380 Å, significantly larger than the PaV capsid at 360 Å. The density distribution of the MGNNV map shows two shells, at radii corresponding to protein, separated by low density. The outer shell is between radii of 154 Å and 192 Å, with a maximum at 168 Å; the inner shell is between radii of 112 Å and 154 Å, with a maximum at 135 Å. The MGNNV map is contoured at a level such that the resultant volume of these two shells is equal to the expected volume of its protein capsid, which is formed by 180 copies of a 37 kDa subunit. It is noteworthy that the interior

boundary of the inner shell (112 Å) of MGNNV and PaV are nearly coincident, and this defines the interior limit of the β-sandwich domain of PaV. Density with a radius of less than 112 Å is probably predominately RNA. The outer shell of MGNNV is composed mainly of the large protrusions located at the quasi-threefold axes. These protrusions are much more prominent than those of PaV. The inner shell is relatively uniform, and the protrusions are separated from it by a void of nearly 12 Å, indicating that the connection between the domains is probably a single, extended polypeptide chain that is not visible in the cryoEM map. The crystal structures of PaV and Flock house virus showed that the protrusions at the highest radius are formed by three two-stranded β-sheets related to each other by quasi-threefold symmetry. These β-sheets are formed by insertions between strands of the canonical viral β-sandwich that forms the contiguous protein shell of the insect nodaviruses, and the strands are twisted together about the quasi-threefold axes to form the surface protrusions. However, the inner and outer regions of the insect nodaviruses display continuous density in contrast to the density gap between the outer and inner shells of protein in MGNNV. Superposition of the cryoEM map of MGNNV with the atomic model of PaV derived from the crystal structure showed relatively good agreement for the region occupied by the contiguous β-sandwich shell in PaV. However, there was poor agreement in the outer radial region. The protrusions at the quasi-threefold axes in MGNNV have much larger volumes than could be accounted for by the three twisted β-sheets at the surface of the PaV structure. Thus, these protrusions must contain more protein that may form individual domains. Moreover, the PaV model could not account for differences in density observed in the MGNNV reconstruction at the icosahedral and quasi-twofold symmetry axes. Significant density exists at the icosahedral two-fold axes between protrusions. However, there is only weak density at the corresponding regions between protrusions related by the quasi-twofold symmetry. This density difference implies different patterns of contacts between protrusions at icosahedral and quasi-twofold axes. Both the cryoEM reconstruction and the crystal structure of PaV revealed that an ordered portion of the viral RNA forms a dodecahedral cage

composed of 30 segments of RNA duplex closely associated with the capsid. In the reconstruction of MGNNV, there is no significant density adjacent to the capsid that can be interpreted as duplex RNA. However, the density at lower radius (112 Å) may be attributed to the cellular RNA randomly packaged inside the capsid and/or portions of the coat protein, e.g. the N-terminal basic-residue-rich segments that are likely to interact with RNA.

Modeling the Domain Structure of MGNNV Capsid Protein

The MGNNV capsid protein shows no significant sequence homology with any known insect nodavirus. We used 3D-PSSM (three-dimensional position-specific scoring matrix) to investigate the folding motif of the MGNNV coat protein.[5] Four structures were identified as comparable with MGNNV with a confidence level of 95%, while the rest of the candidates had a confidence level of 50%. Strikingly, there were no insect nodaviruses with known atomic structure in the top 20 comparable structures. All four structures identified as likely homologues of MGNNV were virus coat proteins that have canonical, uninterrupted viral β-sandwich folds. The first three varied from 189 to 222 aligned residues, and they predicted the fold of MGNNV residues 31 to 235, 49 to 214, and 27 to 213, respectively, as a β-sandwich (Fig. 2). The fourth structure provided prediction for almost the entire MGNNV sequence (residues 24 to 319). The fourth structure was the coat protein of tomato bushy stunt virus (TBSV), which contains two domains: the N-terminal domain (residues 103 to 271) forming a β-sandwich and the contiguous protein capsid, and the C-terminal domain (residues 272 to 387) forming a protrusion at the capsid surface. Sequence comparisons of the coat proteins of several fish nodaviruses showed that the protein could be divided into a conserved region and a variable region. The conserved region comprises residues 83 to 216 and shows pairwise sequence identity of 86% to 96%; the variable region comprises residues 235 to 315, with pairwise sequence identity of 66% to 84%. These results suggest that residues 83 to 216 of MGNNV might form a conserved β-sandwich

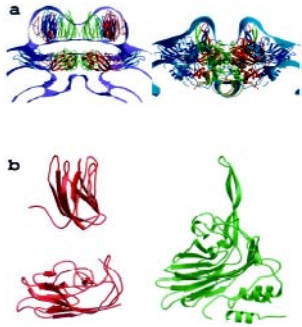


Fig. 2. Cross-sectional views of the particle coat of MGNNV (*left*) and PaV (*right*) defined by the cryoEM density.
(Reprinted from Tang *et al.*, 2002.)

domain that is similar to that in TBSV. Residues 217 to the C-terminal residue 338 would then form a protruding domain that corresponds to the surface protrusions observed in the cryoEM map.

The β-sandwich domain (residues 103 to 271) and the protruding domain (residues 272 to 387) of TBSV were manually fitted into the inner shell and protruding density of the cryoEM map of MGNNV, respectively. The resultant model was subjected to cycles of rigid body refinement against a set of structure amplitudes generated from the cryoEM map. This adjustment ensured that the refinement was carried out against a molecular volume and shape. During the refinement, each domain was defined as an independent rigid group. The final model agreed well with the cryoEM map. Compared to the original TBSV coordinates, the β-sandwich domain was shifted slightly toward the center of the virus particle whereas the protruding domain moved by 24 Å with a rotation of 87°. Such a large change reflects an improved fit to the density but also suggests, like the 3D-PSSM searches, that there may be no relationship between the folds of the outer domains in TBSV and MGNNV and that the TBSV domain is only shifted to optimally occupy the MGNNV density. Nevertheless, the pattern of interactions between the outer domains confirms the $T = 3$ lattice of the capsid. These domains make more extensive contacts at the icosahedral twofold axes than at the quasi-twofold axes, which is consistent with the observation of more significant density at

icosahedral twofold axes than at quasi-twofold axes in the cryoEM map. The model suggests that the C terminus of the coat protein of MGNNV is probably located at the outer surface of the capsid. In contrast, the X-ray structures of insect nodaviruses showed that the C-terminal regions are situated inside the capsid. Presumably, residues around 216 form a polypeptide hinge to connect the two domains. It is likely that the relative position and orientation of the outer domain in MGNNV are different from those in TBSV. The protruding domains in MGNNV are held together around the quasi-threefold axes to form 60 protrusions, whereas the protruding domains in TBSV join at the icosahedral and quasi-twofold axes to form 90 protrusions.

VLPs Compete Against Infection by Native DGNNV

The surface properties of VLPs can be characterized by their ability to compete with infection by the native virus. The first step of infection is attachment to the cell membrane. CPE is then observed when the virus multiplies in the host cells. As a viral structure analogue, VLPs should be able to specifically occupy virus receptor sites and block entry of the virus into the cell. Therefore, occurrence of CPE and attachment to the cell surface were employed to test the structural properties of the DGNNV VLPs. CPE resulting from native virus infection was characterized by the formation of a vacuole in the cytoplasm and the disintegration of cells four days post-infection (p.i.).[6] In the presence of VLPs, cells were protected from invasion by the virus and retained almost complete integrity at four days p.i. As observed in a fluorescent microscope, the cell surface did not bind FITC-labeled virus in the presence of VLPs, while FITC-labeled virus and VLPs alone attached to the cell surface. These results suggest that the VLPs formed in *E. coli* can bind to the surface protein of cell membranes and block early virus-cell recognition and the effective internalization of the virus. The virus receptor on the cell membrane can specifically bind both native virus and VLPs. The outer shell of DGNNV VLPs expressed in *E. coli* is thus structurally indistinguishable from native virus. To our knowledge, this is the first report that the piscine nodavirus capsid protein can form VLPs in *E. coli*. The VLP

expression system in *E. coli* will be used for structural comparison to elucidate the correlation between the size/shape and the RNA content. These mutant VLPs will be employed to study their structure variation by using electron cryomicroscopy and X-ray diffraction. High similarity in structure implies that the VLPs could induce a strong immune response against virus. In addition to their structural and immunological properties, DGNNV VLPs with no virus genome are not infectious, therefore qualifying as a safe and efficient vaccine (Conner et al., 1996; Roy, 1996). Future experiments on VLPs with a variety of mutant RNA1 and RNA2 species will help further understanding of the recognition and binding between the RNAs and capsid protein.

Acknowledgments

The grants were partially supported by the National Science Council, Taiwan (NSC92—2313-B-110-005) and by the University Integration Program from the Ministry of Education (93RA04602).

References

1. Mori KI, Nakai T, Muroga K, *et al.* (1992) Properties of a new virus belonging to Nodaviridae found in larval striped jack (*Pseudocaranx dentex*) with nervous necrosis. *Virology* **187**: 368–371.
2. Munday BL, Kwang J, Moody N. (2002) Betanodavirus infections of teleost fish: a review. *J Fish Dis* **25**: 127–142.
3. Lin CS, Lu MW, Tang L, *et al.* (2001) Characterization of virus-like-particles assembled in a recombinant baculovirus system expressing the capsid protein of a fish nodavirus. *Virology* **290**: 50–58.
4. Lu MW, Lin CS. (2003) Involvement of the terminus of grouper betanodavirus capsid protein in virus-like particle assembly. *Arch Virol* **148**: 345–355.
5. Tang L, Lin CS, Krishna NK, *et al.* (2002) Virus-like particles of a fish nodavirus display a capsid subunit domain organization different from insect nodavirus. *J Virol* **76**: 6370–6375.
6. Lu MW, Liu W, Lin CS. (2003) Infection competition against grouper nervous necrosis virus by virus-like particles produced in *Escherichia coli*. *J Gen Virol* **84**: 1577–1582.

Chapter 17

The Assembly of the Double-Layered Capsids of Phytoreoviruses

Toshihiro Omura, Naoyuki Miyazaki†, Hisashi Naitow‡, R. Holland Cheng§, Tomitake Tsukihara† and Atsushi Nakagawa†*

Introduction

It seems reasonable to assume that a virus — a supramolecule with a variety of biological functions — would have a structure that is advantageous with respect to the performance of these functions. Furthermore, we would expect that the organization of the viral proteins would be genetically economical for the packaging of the viral genome and related proteins within the restricted cavity created by the capsid proteins.

Viruses in the family *Reoviridae*[1] have an inner core with a large interior cavity that contains the double-stranded RNA genome, which consists of 10 to 12 segments, and a transcriptional complex that includes proteins with RNA polymerase, helicase, guanylyltransferase

*Corresponding author. National Agricultural Research Center, 3-1-1 Kannondai, Tsukuba, Ibaraki 305-8666, Japan.
†Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan.
‡Division of Bio-Crystallography Technology, RIKEN Harima Institute, 1-1-1 Kouto Mikazuki, Sayo, Hyogo 679-5148, Japan.
§ Professor of Molecular and Cellular Biology, College of Biological Sciences, University of California, Davis Briggs Hall, Davis, CA 95616, USA.

and transmethylase activities and an RNA-binding protein. The core particle is surrounded by one or two layers of outer capsid proteins. During studies of the structural organization of *Rice dwarf virus* (RDV), a member of the genus *Phytoreovirus* in the family *Reoviridae*, we have identified possible structural mechanisms that allow creation of a large cavity inside a double-layered spherical particle that consists of heterologous proteins with different lattices at the atomic level. The viral particle seems to be created in a genetically economical manner, with the sealing of joints between inner-layer proteins by a second layer of proteins, suggesting the organization of a rigid protein layer that is separate from the interior of the virus and probably protects the interior from the cytoplasmic environment within infected cells.

The Core Capsid Protein of RDV has Conformational Properties that Allow Creation of a Large Cavity Inside the Core

The physical properties of RDV can be summarized as follows. The molecular mass of the virus is 7.5×10^7 Da, and dsRNA accounts for approximately 20% of the total mass. In the inner core there are 12 segments of dsRNA, each of which encodes a single protein. The genome is 25.7 kilo base pairs (kbp) long, with a molecular mass of 1.64×10^7 Da. Each viral particle consists of two concentric layers of proteins. The core of the virus is composed of P3 proteins that encapsidate the 12 segments of dsRNA plus the proteins that are required for transcription. The core is surrounded by the outer capsid layer, and the major component of this layer is the P8 protein.

Studies of the three-dimensional structure, determined at 3.5 Å resolution,[2] revealed the mechanism of assembly of RDV, with its double-layered capsid composed of two kinds of protein. The inner layer of the capsid is composed of 120 copies of P3 protein. It has T = 1 symmetry and, as found in other viruses in *Reoviridae*, it is composed of two icosahedrally independent P3 subunits, P3A and P3B (Fig. 1(a)). In this respect it resembles bluetongue virus (BTV) and orthoreovirus, which are animal reoviruses.[3,4] The outer, second layer of proteins exhibits T = 13*l* icosahedral symmetry[5,6] and is

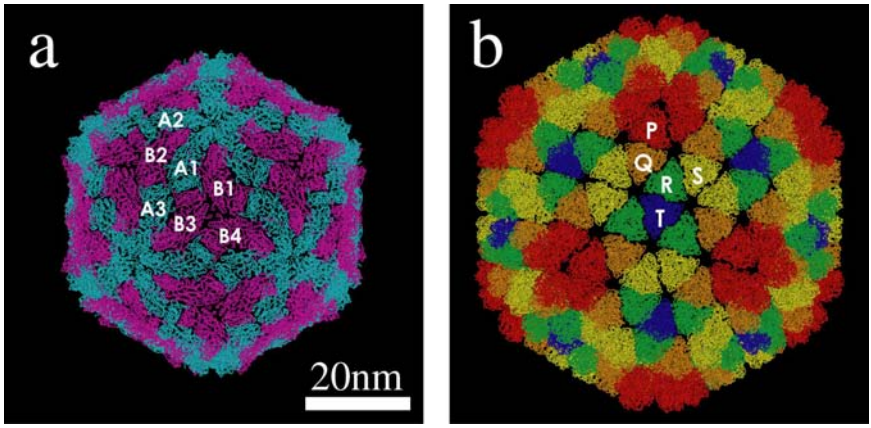


Fig. 1. Structure of *Rice dwarf virus*.[2] **(a)** Cα-trace of the core structure. The icosahedrally independent molecules P3A and P3B are shown in light blue and pink, respectively. Notations of individual P3 proteins are also shown (the letters A and B denote icosahedrally independent molecules). **(b)** Cα-trace of the outer shell of RDV. Each icosahedral asymmetric unit contains 13 copies of P8 proteins, designated P, Q, R, S, and T and colored red, orange, green, yellow, and blue, respectively. The representations were generated using MOLSCRIPT[24] and rendered with RASTER3D.[25]

composed of P8-trimers (Fig. 2(a)), which we have designated P, Q, R, S and T, respectively (Fig. 1(b)).

It is noteworthy that the core capsid protein P3 is very thin (along the "z" axis) varying from 25 to 45 Å in thickness, but it covers a relatively wide area of 65 × 150 Å (Fig. 3(a)). The atomic structures of the innermost first-layer capsid proteins of BTV[3] and orthoreovirus[4] are basically the same as that of P3 of RDV (Figs. 3(b) and 3(c)). These proteins seem to have the ability to generate the spacious cavity within the viral core. However, this type of structure implies that side-by-side binding of the proteins must be very weak because of the small area of the face of each that interacts with others. Our group found that expression of the core capsid protein P3 of RDV in a baculovirus system resulted in the formation of single-layered core-like particles in insect cells in the absence of any other capsid proteins,[7] indicating that P3 has the intrinsic ability to form core particles. Thus, P3 appears to have the conformational properties

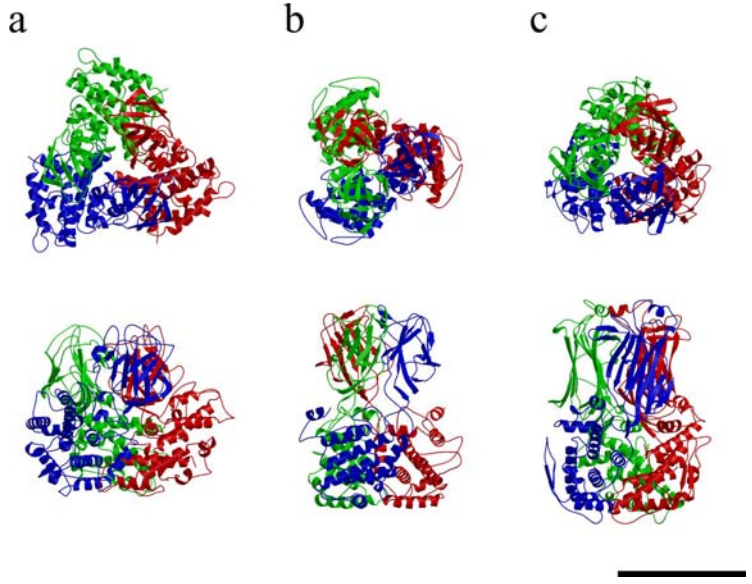


Fig. 2. Comparison of P8 of RDV with VP7 of BTV and VP6 of rotavirus. Top: Trimers of P8 of RDV **(a)**, VP7 of BTV **(b)**, and VP6 of rotavirus **(c)**, as observed from the outside of the viral particle. Individual subunits are colored red, blue, and green, respectively. Bottom: The same images after rotation through 90°.

that allow the bilateral binding of homologous P3 proteins that results in the self-assembly of a spherical particle.

When our group analyzed the structure of P3 by X-ray crystallography,[2] we found that subunits A1 and B1 and their equivalent pairs (see Fig. 1(a)) form the most abundant hydrogen bonds and salt bridges among the possible P3-P3 interfaces. The amino-terminal arm of 30 residues of P3B interacts intimately with P3A (Fig. 4), and this arm seems to stabilize the dimer of the two P3 subunits. Such an atomic structure probably corresponds to the dimers of P3 protein, whose formation was demonstrated biochemically.[8] The formation of the dimeric structure — which can be considered to be the unit piece in the jigsaw puzzle that yields the icosahedral core structure when completed — might initiate the structural assembly of the RDV core particle.

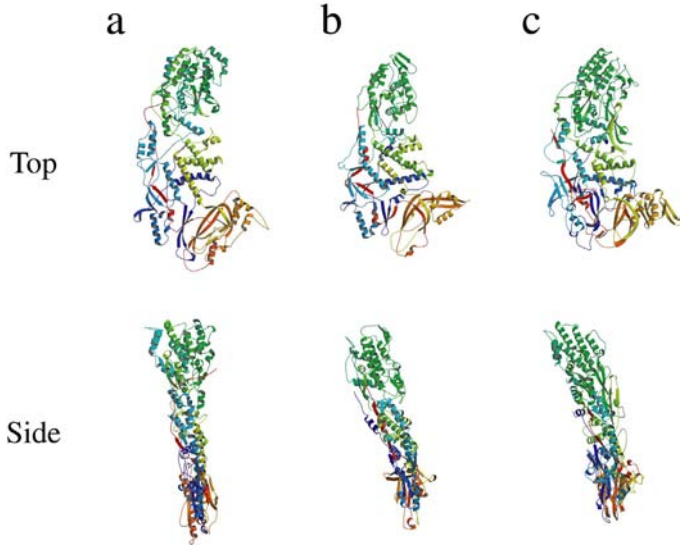


Fig. 3. Comparison of P3 of RDV with VP7 of BTV and λ1 of reovirus. Upper panel: Ribbon representations of P3A (**a**), VP7A (**b**), and λ1A (**c**), viewed from the outside of the core particle. Lower panel: The same images after rotation through 90°.

To identify the role of the amino-terminal region of the P3 protein in the generation of RDV particles, our group produced a series of deletion mutants of P3 with deletions of amino acids 2 through 10, 2 through 29, and 2 through 52 (N10del-P3, N29del-P3, and N52del-P3, respectively). Then we analyzed the stability of the core particles generated by these mutant proteins in the presence of high concentrations of $MgCl_2$.[9]

Core-like particles (CLPs) were observed after expression of N10del-P3 and of N29del-P3, but not after expression of N52del-P3, thus indicating that the amino-terminal region from residue 30 to residue 52 is essential for the self-assembly of core particles. A comparison of the structures of two of the deletion mutants of P3, N29del-P3 and N52del-P3, revealed that N52del-P3 had lost the α-helical domain from residue 32 to residue 51, suggesting that the amino acids in this region might be important for maintenance of

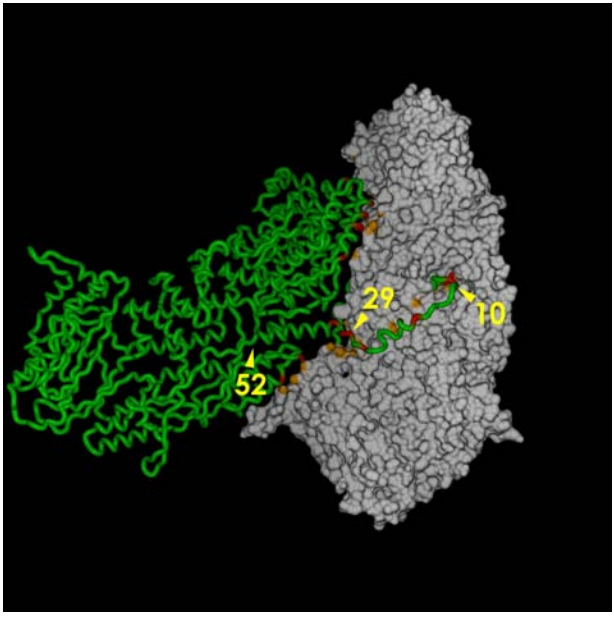


Fig. 4. Structure of the asymmetric dimer composed of P3A and P3B, as shown by a surface representation of the P3A molecule (right, gray) and the Cα-trace of the P3B molecule (left, green), as viewed from the inside of the virus particle. Regions involved in hydrogen bonding or in a salt bridge between the P3 monomers are colored orange (P3A) and red (P3B). The extent of each of three deletions (to residues 10, 29, and 52) is indicated. The representation was produced with the DINO visualization program (http://www.dino3d.org) and rendered with the Povray graphics program (http://www.povray.org) (reproduced with permission).[9]

the overall conformation of P3 and/or for the specific interactions between P3A and P3B that are essential for the self-assembly of CLPs.

In the presence of $MgCl_2$ the stability of CLPs generated from full-length P3 and that of intact RDV particles were the same, and particles of both types were completely dissociated by 2.6 M $MgCl_2$. By contrast, CLPs prepared from N10del-P3 were dissociated by 2.4 M $MgCl_2$ and CLPs prepared from N29del-P3 were even more sensitive to $MgCl_2$. Most of the CLPs prepared from N29del-P3 were dissociated by 2.0 M $MgCl_2$ and all were completely destroyed by treatment with 2.2 M $MgCl_2$. These results indicated that the amino-terminal 29 amino

acids of P3 are important for the stabilization of CLPs at high concentrations of $MgCl_2$. The theoretical interactions between two molecules of P3 in an asymmetric dimer include 11 strong hydrogen bonds and four strong salt bridges within this amino-terminal arm among the 19 strong hydrogen bonds and 12 strong salt bridges involved in the entire interaction between P3s.[7] Thus, the stability of P3A-P3B interactions fell gradually with increased deletion of amino-terminal residues. These results suggest that the reduction in stability of the CLPs obtained with the amino-terminally deleted mutant P3 proteins was caused by loss of the bonds described above.

Calculations of the total energy of interaction for asymmetric dimers in reoviruses, based on their atomic structure, indicated that asymmetric P3 dimers of RDV have a higher energy of interaction (207.3 kcal/mol) than those of VP3 of BTV (134.1 kcal/mol), and the amino-terminal region seems to be responsible for this difference. In the case of N52del-P3 of RDV, which had lost the capacity for self-assembly, the total energy of the P3A-P3B interaction was less than half (92.2 kcal/mol) of that of the native P3-P3 interaction. The relatively high energy of interaction between P3 monomers in P3 dimers in RDV, due to the amino-terminal region of P3, probably allows cores of RDV to form in the absence of other structural proteins. The requirement for an additional protein (VP7) in BTV[10,11] for the construction of CLPs might be due to the lower energy of the interaction between the monomers that correspond to P3 in BTV. In orthoreovirus, the formation of inner core particles by mutant core-capsid protein $\lambda 1$ — from which the amino-terminal 230 amino acids of the wild type have been removed (N230del-ramda1; 99.5 kcal/mol) — might be due to assistance by the $\delta 2$ protein.[12]

Inside the first-layer capsid of RDV there is a 25.7 kbp dsRNA with 12 segments.[13] To our knowledge, this genome is the largest among those of dsRNA viruses studied by X-ray crystallography to date. The cavity also contains P1, a putative RNA-dependent RNA polymerase; P5, a putative guanylyltransferase; and P7, a non-specific nucleic acid-binding protein. It seems reasonable that a large cavity should be required to enclose both dsRNA and the molecules involved in transcription. The generation of strongly interacting dimers that allow the side-by-side binding of the very thin P3 proteins[2] allows creation of this large cavity.

Thus, RDV produces two types of P3 capsid protein that bind tightly to each other and form dimmers, which are the building blocks of the core particle. It seems that only one gene (for P3) is required for the synthesis of the two types of molecule that are needed for creation of the large cavity. This system is rather economical when viewed from a genetic perspective. It will be of interest to examine the events that are responsible for the post-translational discrimination between the two types of P3 molecule in the dimer.

Structural Mechanism that Creates Symmetry Mismatch Between the Inner and Outer Layers and Results in the T = 13*l* Structure of the Outer Layer

The next task in the formation of RDV is the construction of the second layer of the capsid. The double-layered capsid of RDV has different icosahedral lattice symmetries in the inner (T = 1) and outer (T = 13*l*) capsid. In the outer capsid, five icosahedrally unique P8-trimers (designated P-, Q-, R-, S- and T-trimers) are located in different environments on the surface of the inner core particle (Fig. 1). The concept of T number has provided an index for the display of quasi-equivalent lattice systems of viral capsid structures.[14] The structure described above seems to be common to viruses in *Reoviridae*, such as rotavirus (Prasad *et al.*, 1996),[15] BTV[3] and RDV.

As described above, there are two kinds of interaction in the capsid: one type between homologous proteins (P3-P3 and P8-P8) in the intra-layer, with lateral interactions; and one type between heterologous proteins (P3-P8) in the inter-layer, with vertical interactions. Both types of interaction are necessary for the formation of the double-layered capsid.

When studies of the atomic structure of RDV focused on the binding between heterologous proteins P3 and P8, the strongest interactions between the P8-trimer and P3 proteins were found in the three-fold symmetry region of the viral core. Total energies of interaction were 53.7, 110.3, 68.5, 49.3 and 153.3 kcal/mol for the P-, Q-, R-, S- and T- trimers, respectively (Fig. 1(b)). This result supported the observation, made by cryo-electron microscopy,[16] that only the T-trimers remain on the surface of the core after treatment of intact particles

with 0.8 M $MgCl_2$. The atomic-structural analysis indicated that the strongest interaction between the T-trimer and P3B at the three-fold axis involved electrostatic interactions, hydrogen bonds, and surface complementarity (Fig. 5). This intimate interaction seemed to determine the orientation of the margin of the T-trimer relative to the

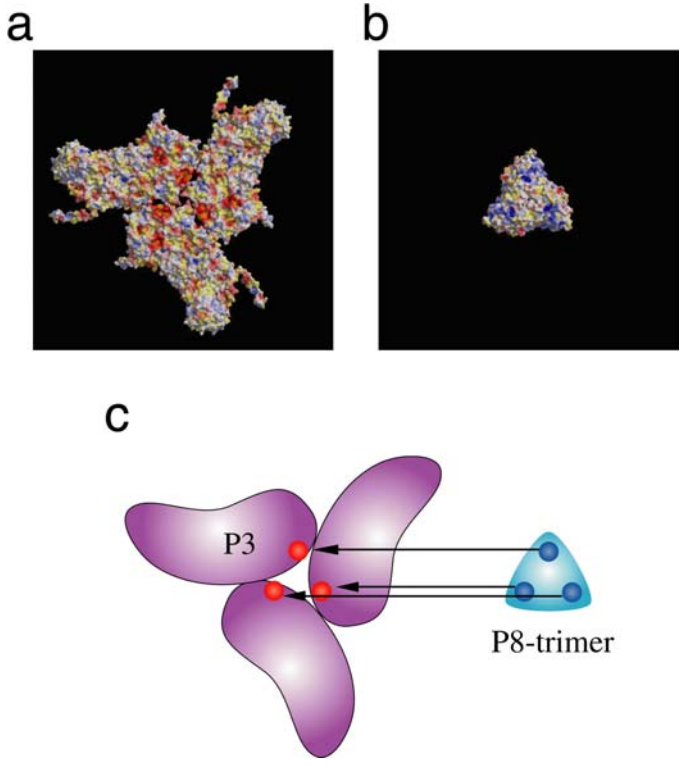


Fig. 5. Interactions between the core and the outer shell.[2] **(a)** Electrostatic potential on the surface of the core at the icosahedral three-fold axis, as viewed from the interface between the core and the outer shell. The map of electrostatic potential was calculated by the program used for the eF-site database.[26] **(b)** Electrostatic potential on a P8-trimer, as viewed from the interface between the core and the outer shell. **(c)** Schematic representation of the interaction between the inner shell and a P8-trimer at an icosahedral three-fold axis. Positively charged patches on the P8-trimer (shown in blue) correspond closely to negatively charged regions on the surface of the core (shown in red) at the icosahedral three-fold axis.

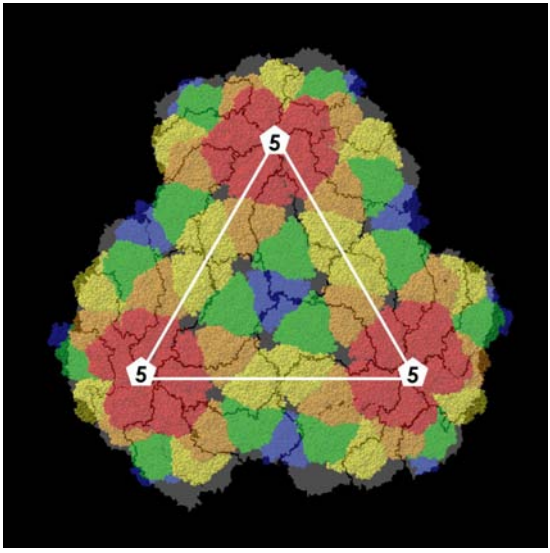


Fig. 6. Superimposed representations of components of the double-shelled structure around an icosahedral three-fold axis. Five icosahedrally unique P8-trimers sit on different environments on the surface of the core particle. Spaces between inner core proteins are well covered by outer P8-trimers.

surface of the core. When we examine the structure of RDV in this context, we see that the edges of the P8-trimer are not parallel to the line that can be drawn between adjacent icosahedral five-fold axes (Fig. 6). This discrepancy suggests the absence of any regular correspondence between the core proteins and the outer capsid proteins, which might result, for example, in a T = 9 structure. This structural feature would determine subsequent binding of P8-trimers and result in a T = 13 structure, which has been hard to interpret, particularly from a biochemical perspective, because some of the P8-trimers should span the margins of icosahedral particles.

Biochemical evidence to support of the mechanism of assembly of double-layered capsids with disparate icosahedral lattices, as described above, was obtained from studies of chimeric virus-like particles with components derived from different viral strains, which allowed discrimination among the respective proteins in assembled particles either

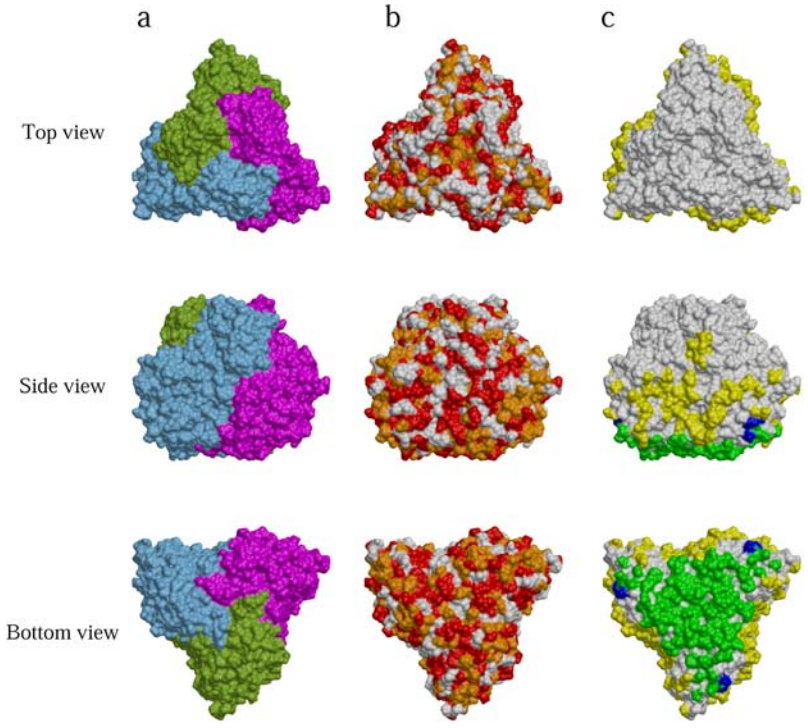


Fig. 7. Distribution of variations in the sequence of P8 between RDV and RGDV.[21] Surface representations are shown of P8-trimers viewed from the top, side and bottom. **(a)** The three P8 subunits in a P8-trimer are colored blue, magenta and green. **(b)** Regions of identical sequence in P8 of RDV and RGDV are colored red, regions of similar sequence are colored orange, and regions of non-similar sequence are colored white. **(c)** The regions that make contact with neighboring P8-trimers and the P3 layer are highlighted. The surfaces in contact with neighboring P8-trimers, with the P3 layer, and with both neighboring P8-trimers and the P3 layer are colored yellow, green and blue, respectively. Representations were produced with DINO (http://www.dino3d.org).

serologically or biochemically. Three-dimensional homology mapping of P8 proteins in RDV and RGDV (Fig. 7) indicated that the amino acid residues at the interface between the T-trimer and the core particle — where heterologous P3 core and P8 outer capsid proteins interact — were more strongly conserved than those that on the upper

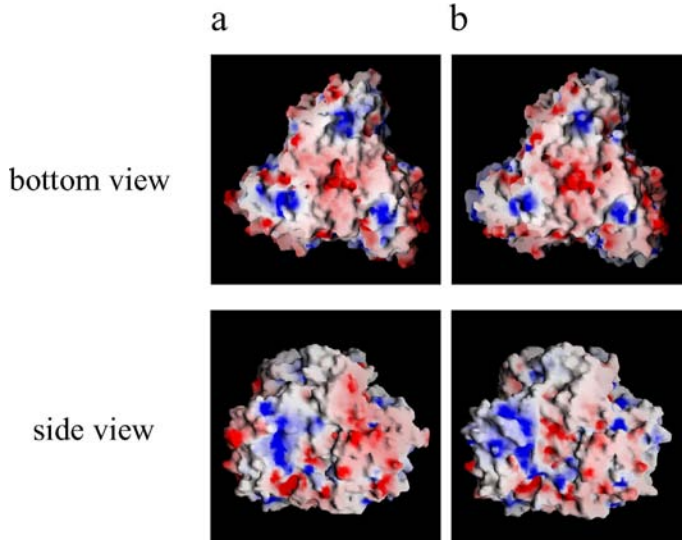


Fig. 8. Distribution of electrostatic potential on a P8-trimer, as viewed from the interface between the core and the outer shell and between P8-trimers.[21] Positively charged regions are colored blue and negatively charged regions are colored red. The atomic model of the P8-trimer of RGDV (b) was based on that of the P8-trimer of RDV (a); figures were prepared with GRASP.[27]

surface of the P8 protein, namely, on the surface of the virus. Furthermore, the structure of the P8-trimer of RGDV, constructed on the basis of the atomic structure of the P8-trimer of RDV, had a similar distribution of electrostatic potential to that of the bottom of the P8-trimer of RDV (Fig. 8, top panels), which is the region that is assumed to play a critical role in the interaction between the P8-trimer and the three-fold axes of core particles.[2] The two kinds of similarity on the lower part of the P8-trimer, which binds to the core particle, suggested that it might be possible for P8-trimers of RGDV to bind at the three-fold axes of RDV core particles, i.e. via the vertical interactions of heterologous proteins. The validity of this suggestion was confirmed by the demonstration that treatment with 0.4 M $MgCl_2$ of core particles prepared from chimeric virus-like particles, obtained by co-expression of P3 of RDV and P8 of RGDV, contained relatively small amounts of RGDV P8 protein, which were resistant to treatment

with higher concentrations of $MgCl_2$, as reported to be the case for T-trimers that bind tightly at the three-fold axes of the core.[16] This result supports the hypothesis that a P8-trimer binds to the region of three-fold symmetry of the viral core at the onset of second-layer assembly, and its orientation determines the orientation of the P8-trimers that bind subsequently.

The studies described above showed that both the presence of regions of conserved amino acids that interact with neighboring capsid proteins and the similar distributions of electrostatic potential on the proteins are required for binding heterologous proteins. Furthermore, these observations confirmed, by biochemical methods, the mechanism of assembly of the double-layered capsid with disparate icosahedral lattices, which was suggested by structural analysis. Such novel architecture might be representative of other layers of multiple proteins that form cavities with functional molecules inside them.

Side-by-side Interactions Among P8-trimers: a Mechanism for Formation of the Second Layer of the Capsid

Once the orientation of a T-trimer relative to the core has been established by specific electrostatic interactions between a P8-trimer and the core along the three-fold axis, the icosahedral T = 13*l* arrangement is probably generated by subsequent interactions between P8-trimers and the core, as well as by interactions among the P8-trimers themselves. Indeed, two-dimensional crystallographic studies of P8 proteins have shown that the P8-trimer has structural features that facilitate side-by-side binding of P8-trimers.[17]

X-ray crystallographic studies of intact RDV[2] revealed the atomic details of this biochemical phenomenon. The electrostatic potential at the margin of each P8-trimer has obvious patches of positive and negative charge, allowing clear electrostatic complementarity between adjacent subunits (Fig. 9). Shape complementarities[18] among trimers (SC = 0.26 to 0.49) and interactions between individual P8 proteins might also stabilize the interactions between adjacent P8-trimers. In particular, a close interaction between two P8-trimers was evident at

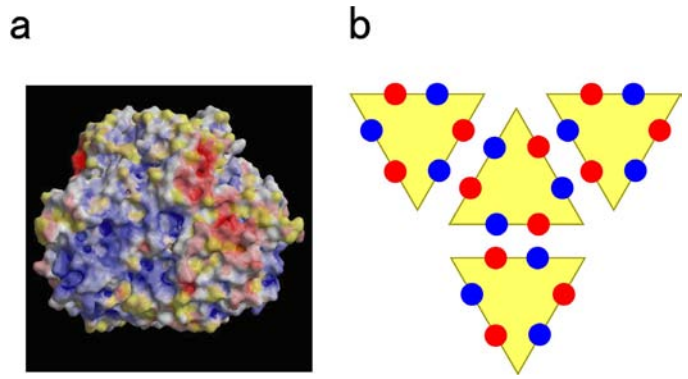


Fig. 9. Interactions among P8-trimers.[2] **(a)** Surface electrostatic potential on a P8-trimer, as viewed from the interface between two P8-trimers. Positively charged regions are colored blue and negatively charged regions are colored red. Charged regions are clustered, suggesting that they might complement the distribution of charges on neighboring P8-trimers. The electrostatic potential map was calculated by the program used for the eF-site. **(b)** Schematic representation of side-by-side interactions among P8-trimers. Positively charged patches (shown in blue) and negatively charged patches (shown in red) make electrostatic pairs at the P8-P8 interfaces.

amino acid residues 274 through 276. This region includes a unique amino acid sequence, namely, a glycine triplet (GGG). The second and third glycine residues are especially close to corresponding residues related by icosahedral two-fold symmetry. A bulky side chain at this position, even a methyl group, might result in steric hindrance of the interactions between P8-trimers. This triplet is semiconserved in viruses that belong to the genus *Phytoreovirus*, such as RGDV[19] and *Wound tumor virus* (WTV),[20] which have SGG and AGG triplets, respectively. Thus, the bilateral features in the atomic structure of P8-trimers might be significant for the interactions that allow self-assembly of the viral outer capsid.

A transcapsidation study was performed to confirm this hypothesis.[21] Three-dimensional homology mapping of P8-trimers indicated that the amino acid residues involved in the interaction between the edges of P8-trimers are strongly conserved (Fig. 7), as is the case for

the amino acid residues in the bottom region, which interact with the surface of the core particle, as described above. Furthermore, the electrostatic potential at the edge of P8-trimers of RGDV, which might similarly play an important role in the interactions between P8-trimers in RDV,[2] revealed obvious patches of positive and negative charge, which would allow clear electrostatic complementarity among adjacent subunits (Fig. 8, side view). This electrostatic complementarity would allow a second outer layer to form with a different and heterologous viral protein[22] because of similar amino acid sequences and electrostatic similarities, in spite of the fact that the homology at the amino acid level between the heterologous P8 proteins from RDV and RGDV is only 52%. Thus, binding of neighboring P8-trimers to the T-trimer via the mutual binding of P8-trimers allowed P8 to polymerize and cover the total surface of the core particle in a mismatched manner, with a T = 13*l* structure.

When sitting of P8-trimers on the P3 core capsid proteins was analyzed, it was clear that all the junctions formed by the organization of P3 protein monomers were covered by P8-trimers (Figs. 1 and 6). Thus, it seems likely that the role of the T = 13*l* structure of the second and outer layer, found in the viruses that belong to *Reoviridae*, might be to protect the interior portion of the virus from attack by the agents in the host cell's cytoplasm, such as ions and free radicals, for example. If this is the case, the mismatches in the T = 13*l* structure, in which P8-trimers bind at environmentally different sites on the inner surface of the capsid, provide a structure that, while hard to understand from a biochemical perspective, apparently represents a sophisticated defense by the virus against, for example, attack by its host.

The Rigid Structural Unit of the Outer Capsid, the P8-trimer

Unlike the first-layer capsid protein P3, which is thin and flat, the P8-trimer, which is the subunit that forms the second layer, appears to be granular in shape. When P8 proteins form trimers, each P8 protein wraps around another in a right-handed manner and subunits "swap"

domains with adjacent subunits via strong interactions that yield a single trimeric unit. The overall shape of the P8-trimer is prismatic, with sides of 90 Å and a height of 75 Å (Fig. 2). This trimeric structure is conserved among viruses that belong to *Reoviridae* (Fig. 2).

The structure of P8-trimers seems to contradict the idea that a capsid protein should be thin and flat so that large amounts of genome and related proteins can be packaged inside a cavity, as proposed for the P3 protein of RDV above. Nonetheless, it is reasonable to postulate that the second layer is made up of proteins that form a structure wherein each trimer seals the spaces at the junctions between individual first-layer proteins. The trimeric structure would be very suitable for a unit that attaches to a three-fold axis at which three molecules of P3 make a triangle joint. Furthermore, a triangular structure would be most suitable for lateral organization on an icosahedral particle. A trimeric structure provides an easy basis for construction of a triangular structure by protein molecules. It would be difficult, for example, for a single protein molecule to create a triangular structure and then for such structures to bind to one another side by side. Furthermore, triangular structures composed of trimers have the advantage that each edge is identical in terms of both conformation and electrostatic charge (Fig. 9). Thus, each edge of each trimer always faces a complementary arrangement when two trimers are adjacent to each other. These features might explain why all second layers of reoviruses examined to date are based on trimers.

It is interesting to note that the height (z axis) of the protein in the second layer of the reoviruses that have been examined is similar, namely, 97 Å in rotavirus,[23] 90 Å in BTV[3] and 75 Å in RDV. We might expect this layer to be thinner from an economical perspective, as exemplified by the first-layer P3 protein and mentioned above. However, it is possible that the second-layer protein needs to be of a certain height to allow tight binding of the trimers that are the building blocks of the second layer. If the intra-P8 binding in the trimer is not as tight as that of inter-P8 binding, various types of building block might be formed and make it difficult to generate an ordered lateral arrangement of P8-trimers in the second layer of the reoviruses. The individual components of the trimer subunits of reoviruses appear to

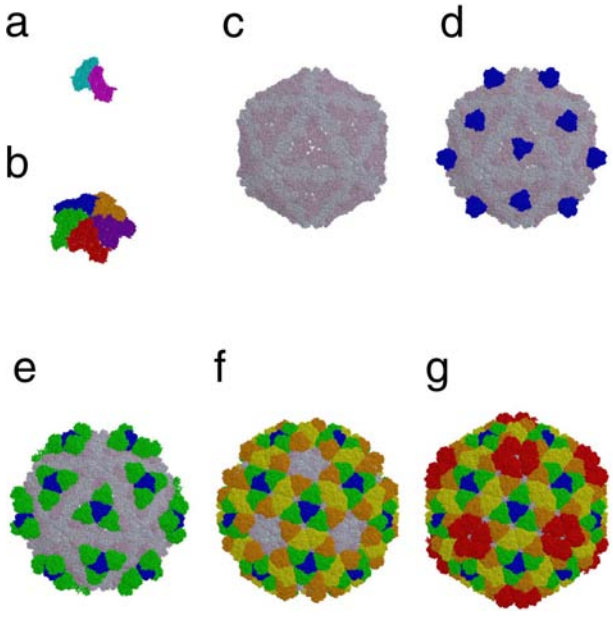


Fig. 10. Proposed mechanism for the assembly of RDV.[2] On the basis of the strengths of interactions among the various subunits, we propose the following sequence of events. **(a)** Insertion of the amino-terminal arm of P3B into P3A initiates the assembly of a P3 dimer. This P3A-P3B dimer acts as the unit piece in the jigsaw puzzle. **(b)** Five dimers of P3 protein aggregate around an icosahedral five-fold axis. **(c)** The pentameric aggregates of P3 assemble to form the core structure of the RDV particle. **(d)** Trimers of P8 act as unit pieces for the assembly of the outer layer and these trimers attach to the icosahedral three-fold axis at the T-site first. Orientation of the T-trimer of P8 proteins on the surface of the core at the icosahedral three-fold axis is defined by electrostatic complementarity. **(e)** R-trimers of the P8 protein then attach via interactions with the inner shell and with the T-trimers. **(f)** Q-trimers and S-trimers attach to the core surface. **(g)** At the final stage of viral assembly, P-trimers attach at the icosahedral five-fold axes to generate the complete virus particle.

twist around each other, with tight binding at their intermolecular junctions (Fig. 2). The binding energies are very high for all the reoviruses for which such data are available: 175.6 kcal/mol for rotavirus; 169.6 kcal/mol for BTV; and 179.8 kcal/mol for RDV. The strong

interactions between components of the trimers in the outer layer of the capsid of viruses in *Reoviridae* stabilize the building blocks of the second layer of the capsid. The increased height (as compared to the inner layer) of these building blocks may be the result of the need for these strong stabilizing interactions.

A virus with a big cavity inside a spherical shell can be generated by a small variety of proteins — that is to say, in a genetically economical manner — either by use of "flat" proteins, as described above for P3 of RDV, or by construction of a capsid layer with a large T-number. The disadvantage in the former case is the weakness of inter-protein binding because of the limited area (margins) of the protein that is available for such binding; the disadvantage in the latter case is the weakness of interactions because so many molecules bind to one another. The weak interactions of the core capsid layer of RDV seem to be reinforced by the well-fitting outer layer, composed of the P8 protein. Conversely, the well-ordered outer layer seems to be supported by the inner layer, which provides a scaffold to which the P8 T-trimers can bind tightly. Thus, the first-layer inner capsid and the second-layer outer capsid proteins compensate for each other's respective weaknesses and form a rigid double-layered shell that creates a cavity that is large enough for the genome and other essential proteins.

References

1. Mertens PPC, Duncan R, Attoui H, Dermody TS. (2005) Family *Reoviridae*. In CM Fauquet, MA Mayao, J Maniloff, *et al.* (eds), *Virus Taxonomy*. Elsevier Academic Press.
2. Nakagawa A, Miyazaki N, Taka J, *et al.* (2003) Atomic structure of a Rice dwarf virus reveals the self-assembly mechanism of the component proteins. *Structure* **11**: 1227–1238.
3. Grimes JM, Burroughs JN, Gouet P, *et al.* (1998) The atomic structure of the bluetongue virus core. *Nature* **395**: 470–478.
4. Reinisch KM, Nibert ML, Harrison SC. (2000) Structure of the reovirus core at 3.6 Å resolution. *Nature* **404**: 960–967.
5. Naitow H, Morimoto Y, Mizuno H, *et al.* (1999) A low-resolution structure of Rice dwarf virus determined by *ab initio* phasing. *Acta Crystallogr D Biol Crystallogr* **55** (Pt. 1): 77–84.

6. Zhou ZH, Baker ML, Jiang W, *et al.* (2001) Electron cryomicroscopy and bioinformatics suggest protein-fold models for rice dwarf virus. *Nature Struct Biol* **8**: 868–873.
7. Hagiwara K, Higashi T, Namba K, *et al.* (2003) Assembly of single-shelled cores and double-shelled virus-like particles after baculovirus expression of major structural proteins P3, P7 and P8 of Rice dwarf virus. *J Gen Virol* **84**: 981–984.
8. Ueda S, Masuta C, Uyeda I. (1997) Hypothesis on particle structure and assembly of rice dwarf phytoreovirus: interactions among multiple structural proteins. *J Gen Virol* **78** (Pt 12): 3135–3140.
9. Hagiwara K, Higashi T, Miyazaki N, *et al.* (2004) The amino-terminal region of major capsid protein P3 is essential for self-assembly of single-shelled core-like particles of Rice dwarf virus. *J Virol* **78**: 3145–3149.
10. Hewat EA, Booth TF, Loudon PT, Roy P. (1992) Three-dimensional reconstruction of baculovirus expressed bluetongue virus core-like particles by cryo-electron microscopy. *Virology* **189**: 10–20.
11. Loudon PT, Hirasawa T, Oldfield S, *et al.* (1991) Expression of the outer capsid protein VP5 of two bluetongue viruses, and synthesis of chimeric double-shelled virus-like particles using combinations of recombinant baculoviruses. *Virology* **182**: 793–801.
12. Kim J, Zhang X, Centonze VE, *et al.* (2003) The hydrophilic amino-terminal arm of reovirus core shell protein lambda 1 is dispensable for particle assembly. *J Virol* **76**: 12211–12222.
13. Omura T, Yan J. (1999) Role of outer capsid proteins in transmission of phytoreovirus by insect vectors. *Advances in Virus Research* **54**: 15–43.
14. Caspar DL, Klug A. (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* **27**: 1–24.
15. Prasad BV, Rothnagel R, Zeng CQ, *et al.* (1996) Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus. *Nature* **382**: 471–473.
16. Wu B, Hammar L, Xing L, *et al.* (2000) Phytoreovirus T = 1 core plays critical roles in organizing the outer capsid of T = 13 quasi-equivalence. *Virology* **271**: 18–25.
17. Zhu Y, Hemmings AM, Iwasaki K, *et al.* (1997) Details of the arrangement of the outer capsid of rice dwarf phytoreovirus, as visualized by two-dimensional crystallography. *J Virol* **71**: 8899–8901.
18. Lawrence MC, Colman PM. (1993) Shape complementarity at protein/protein interfaces. *J Mol Biol* **234**: 946–950.

19. Noda H, Ishikawa K, Hibino H, *et al.* (1991) Nucleotide sequences of genome segments S8, encoding a capsid protein, and S10, encoding a 36K protein, of rice gall dwarf virus. *J Gen Virol* **72** (Pt 11): 2837–2842.
20. Xu Z, Anzola JV, Nalin CM, Nuss DL. (1989) The 3′ terminal sequence of a Wound tumor virus transcript can influence conformational and functional properties associated with the 5′-terminus. *Virology* **170**: 511–522.
21. Miyazaki N, Hagiwara K, Naitow H, *et al.* (2005) Transcapsidation and conserved interactions of two major structural proteins of a pair of phytoreoviruses confirm the mechanism of assembly of the outer capsid layer. *J Mol Biol* **345**: 229–237.
22. Takahashi Y, Tomiyama M, Hibino H, Omura T. (1994) Conserved primary structures in core capsid proteins and reassembly of core particles and outer capsids between rice gall dwarf and rice dwarf phytoreoviruses. *J Gen Virol* **75** (Pt 2): 269–275.
23. Mathieu M, Petitpas I, Navaza J, *et al.* (2001) Atomic structure of the major capsid protein of rotavirus: implications for the architecture of the virion. *EMBO J* **20**: 1485–1497.
24. Kraulis PJ. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* **24**: 946–950.
25. Merrit E, Murphy MEP. (1994) Raster 3D version 2.0, a program for photorealistic molecular graphics. *Acta Crystallogr* **D50**: 869–873.
26. Kinoshita K, Furui J, Nakamura H. (2001) Identification of protein functions from a molecular surface database, eF-site. *J Struct Genomics* **2**: 9–22.
27. Nicholls A, Bharadwaj R, Honig B. (1993) GRASP-graphical representation and analysis of surface properties. *Biophys J* **64**: A166.
28. Omura T, Ishikawa K, Hirano H, *et al.* (1989) The outer capsid protein of rice dwarf virus is encoded by genome segment S8. *J Gen Virol* **70** (Pt 10): 2759–2764.

Chapter 18

Structure and Assembly of Human Herpesviruses: New Insights From Cryo-Electron Microscopy and Tomography

Z. Hong Zhou[*,†,‡] *and Pierrette Lo*[§]

The *Herpesviridae* family is one of the largest and most genetically and structurally complex families of viruses. Eight human herpesviruses have been identified, which cause ailments ranging from cold sores to blindness to cancers. In this chapter, we review advances in cryo-electron microscopy and tomography studies of herpesvirus structure and assembly, focusing on comparisons of representative human herpesviruses in the alpha-, beta- and gammaherpesvirus subfamilies. These herpesviruses share a multilayer architecture with a pleomorphic glycoprotein-containing lipid envelope and a proteinaceous tegument compartment that encloses an icosahedral capsid with a bacteriophage-like portal and a tightly packed double-stranded DNA genome. Capsid structure is highly conserved, with the exception of capsid size, which is related to the varying

*Corresponding author.
[†]Department of Pathology and Laboratory Medicine, University of Texas Medical School at Houston, Houston, TX 77030, USA.
[‡]Present address: Department of Microbiology, Immunology and Molecular Genetics, David Geffen School of Medicine, and the California NanoSystems Institute, University of California at Los Angeles, Los Angeles, CA 90095, USA. E-mail: Hong.Zhou@ucla.edu
[§]Present address: Texas Heart Institute, PO Box 20345, MC3-116, Houston, TX 77225, USA. E-mail: plo@heart.thi.tmc.edu.

genome size of different herpesviruses. In contrast, the outer components exhibit remarkable structural and organizational diversity, including the inner tegument proteins interacting directly with the capsid. These structural results suggests that herpesviruses have evolved their tegument and envelope substantially while adapting to their specific host cellular environment in order to gain optimal fitness and confer unique characteristics of different herpes infections.

Introduction

Herpesviruses are among the largest and most ubiquitous viruses. The family *Herpesviridae* is extensive; in fact, almost every animal species examined can be infected with at least one herpesvirus.[1] Although they vary in host range, genome size and molecular composition, all herpesviruses share a common virion structure consisting of four basic elements: a double-stranded DNA (dsDNA) core, an icosahedral capsid shell, a proteinaceous tegument compartment, and a glycoprotein-decorated lipid envelope. The core of the mature virion consists of dsDNA arranged as closely packed shells.[2] The capsid is a rigid icosahedral protein shell, 1200–1300 Å in diameter, which encloses and protects the dsDNA core. The tegument is a poorly defined, asymmetric layer of host and viral proteins between the capsid and the envelope. It varies in thickness and distribution around the capsid, although some of its lower proteins are anchored to the capsid. The envelope is a host-derived lipid bilayer containing spikes of viral glycoprotein. The entire virion varies in diameter from 1400 to 3000 Å, depending on the thickness of the tegument and the integrity of the envelope.

Herpesviruses have been classified into three subfamilies — alpha-, beta- and gammaherpesviruses — based on shared biological properties, as exemplified in Table 1 by the eight known human herpesviruses.[3] A small number of them have also been classified into genera based on DNA sequence homology, genome arrangement and related proteins. The alphaherpesviruses have a variable host range, a short reproductive cycle and rapid spread in culture. They establish latent infection primarily in neurons. Human pathogens in this subfamily include the human pathogens herpes simplex virus types 1 and 2 (HSV-1/2 or HHV-1/2)

Table 1. Human Herpesviruses and their Associated Diseases

Virus	Subfamily	Genome Length (kb)	Preferred Host Cells	Associated Diseases
HHV-1 (HSV-1)	α	152	Neuroectodermal	Cold sores
HHV-2 (HSV-2)	α	152	Neuroectodermal	Genital herpes
HHV-3 (VZV)	α	125	Neuroectodermal	Chicken pox; shingles
HHV-4 (EBV)	γ	172	Lymphohematopoietic; ectodermal	Infectious mononucleosis; Burkitt's lymphoma; nasopharyngeal carcinoma
HHV-5 (HCMV)	β	230	Mesodermal, incl. lymphohematopoietic	Infectious mononucleosis; sialoadenitis
HHV-6	β	160	Lymphohematopoietic	Exanthem subitum; infectious mononucleosis; Kikuchi's lymphadenitis; infantile febrile seizures
HHV-7	β	145	Lymphohematopoietic	Nonspecific lymphadenitis
HHV-8 (KSHV)	γ	165	Lymphocytic; fibrohistiocytic, incl. endothelial cells	Kaposi's sarcoma; serosa-associated lymphoma

[1]Additional diseases will occur in immunodeficient patients. HHV, human herpesvirus; HSV, herpes simplex virus; VZV, Varicella-Zoster virus; EBV, Epstein-Barr virus; HCMV, human cytomegalovirus; KSHV, Kaposi's sarcoma-associated herpesvirus.

and Varicella-Zoster virus (VZV or HHV-3), which cause cold sores, genital herpes and chickenpox, respectively. The betaherpesviruses have a more restricted host range, longer reproductive cycle and slower growth in culture. The virus can remain latent in secretory

glands, lymphocytes and other tissues. Human cytomegalovirus (HCMV or HHV-5), a leading cause of birth defects, and the human herpesvirus types 6 and 7 (HHV-6, HHV-7), are members of this subfamily. The two known human pathogens of the gammaherpesvirus subfamily, Epstein-Barr virus (EBV or HHV-4) and Kaposi's sarcoma-associated herpesvirus (KSHV or HHV-8), are both associated with lymphomas and other cancers. They have a more restricted host range than viruses of the other two subfamilies. Table 2 lists the major virion-associated proteins of representative human herpesviruses from the three subfamilies.

In this chapter, we will first provide a general account of the current understanding of herpesvirus lytic replication and capsid assembly pathway. We will then review recent structural studies by focusing on three human herpesviruses representative of the three subfamilies: the alphaherpesvirus HSV-1, the betaherpesvirus HCMV, and the gammaherpesvirus KSHV.

Lytic Herpesvirus Infection and Viral Assembly

Our current understanding of the general aspects of the herpesvirus lifecycle can be traced back to earlier studies of the prototypical herpesvirus, HSV-1, as well as similar ultrastructural investigations of other herpesviruses.[4–7] Many of these studies were derived from electron microscope observations of the morphology of viral particles in thin-sectioned, virus-infected cells.

Herpesvirus infection begins with the attachment of virus particles to glycoprotein receptors on the surface of the host cell (Fig. 1). Several cellular receptors for herpesvirus attachment have been identified, including CD21 for EBV, CD46 for HHV-6, and CD4 as a component of HHV-7 binding.[8–11] HCMV binds to a 30 kDa glycoprotein cell membrane receptor, and subsequent virus entry is mediated by HCMV glycoprotein B (gpUL55).[12,13] VZV endocytosis is facilitated by a specific gE:gI Fc receptor that resembles the cellular Fc-γRII.[14] The integrin $\alpha_3\beta_1$ appears to have a role in KSHV endocytosis.[15] Upon binding to the receptor, the viral envelope fuses with either the host membrane or the membrane of an endocytic vesicle, leading to

Table 2. Major Virion-associated Proteins of HSV-1, HCMV, and KSHV

Location	Protein	HSV-1	HCMV	KSHV	Structural or Functional Roles
Capsid shell	MCP	UL19	UL86	ORF25	Forms pentons, hexons
	mCP	UL18	UL85	ORF26	Triplex dimer protein
	mC-BP	UL38	UL46	ORF62	Triplex monomer protein
	SCP	UL35	UL48.5	ORF65	Binds the upper domains of hexons
Portal	Portal protein	UL6	UL104	ORF43	Portal protein
Inside capsid	Protease	UL26	UL80	ORF17	Maturation protease
	Scaffold	UL26.5	UL80.5	ORF17.5	Scaffolding protein
Tegument layer	Tegument protein	UL36	UL48	ORF64	HMWP, binds capsid
	Tegument protein	UL37	UL47	ORF63	Binds HMWP
	Tegument protein	UL25	UL77	ORF19	Binds capsid vertices
	Tegument protein	UL34	UL50	ORF67	Capsid-docking protein on nuclear lamina

(Continued)

Table 2. (*Continued*)

Location	Protein	HSV-1	HCMV	KSHV	Structural or Functional Roles
	Tegument protein	UL24	UL76	ORF20	Fusion protein
	Tegument protein	UL23	Unknown	ORF21	Thymidine kinase
	Tegument protein	No homologue	No homologue	ORF45	IRF7-binding protein homologue
	Tegument protein	No homologue	No homologue	ORF75c	Unknown
Envelope glycoproteins	gB	UL27	UL55	ORF8	Major envelope protein, binds cellular integrin receptor
	gH	UL22	UL75	ORF22	Second most abundant; forms gH/gL/gO
	gp150	No homologue	No homologue	M7	Binds heparin sulfate; conserved among gammaherpesviruses

MCP, major capsid protein; mCP, minor capsid protein; mC-BP, minor capsid binding protein; SCP, smallest capsid protein; HMWP, high-molecular-weight protein; gB/gH/gL/gO, glycoproteins B/H/L/O; IRF, interferon regulatory factor.

receptor-mediated release or endocytosis of a partially tegumented capsid, respectively. The cytoskeletal network, including the dynein and dynactin components, is used to transport capsids to the nuclear pores.[16] The DNA is injected into the nucleus, while the empty capsids remain outside.

In the nucleus, the DNA circularizes and one of two pathways, i.e. latent or lytic replications, is initiated. The viral genome can be maintained as a circular episome and undergo latent replication without producing infectious progeny. It is through this pathway that the genes for cellular transformation are expressed. A latent virus can be reactivated by various factors, such as excessive exposure to ultraviolet light (in the case of HSV-1/2 and VZV), endotoxins, plant agglutinins, steroids, tumor-promoting chemicals, other infections, and increased cytokine production (as in the case of EBV, HHV-6/7 and, in part, HCMV and KSHV). Once reactivated, persistent viral activity can be supported by decreased immune responsiveness resulting from stress or other factors. Most of these factors can trigger transcriptional activation of the lytic switch, ORF50/Rta, resulting in a switch from latent to lytic replication. This pathway leads to the production of infectious virions and the eventual death of the host cell.

During lytic replication, viral DNA is synthesized as a concatemer using the "rolling circle" method. Meanwhile, capsid assembly begins with the formation of a spherical, scaffold-containing procapsid that is similar to the prohead of bacteriophages.[17–21] Cleavage of scaffolding proteins by the viral protease results in a stable, angular form called the B-capsid. The cleaved scaffold proteins are released from the capsid at the same time that DNA is cleaved to genome length and packaged into the mature C-capsid. C-capsids ultimately acquire the tegument and envelope and are released from the cell to begin the infection cycle again. A third form called the A-capsid is an empty shell resulting from aberrant processing of DNA and/or scaffold protein. The entire process by which HSV-1 procapsids mature into angular capsids has been captured by time-lapse cryo-electron microscopy.[22] The procapsid passes through several continuous intermediate stages of maturation on its way to becoming an angular, mature capsid (Fig. 2), through spontaneous, yet poorly understood, pathways.[23]

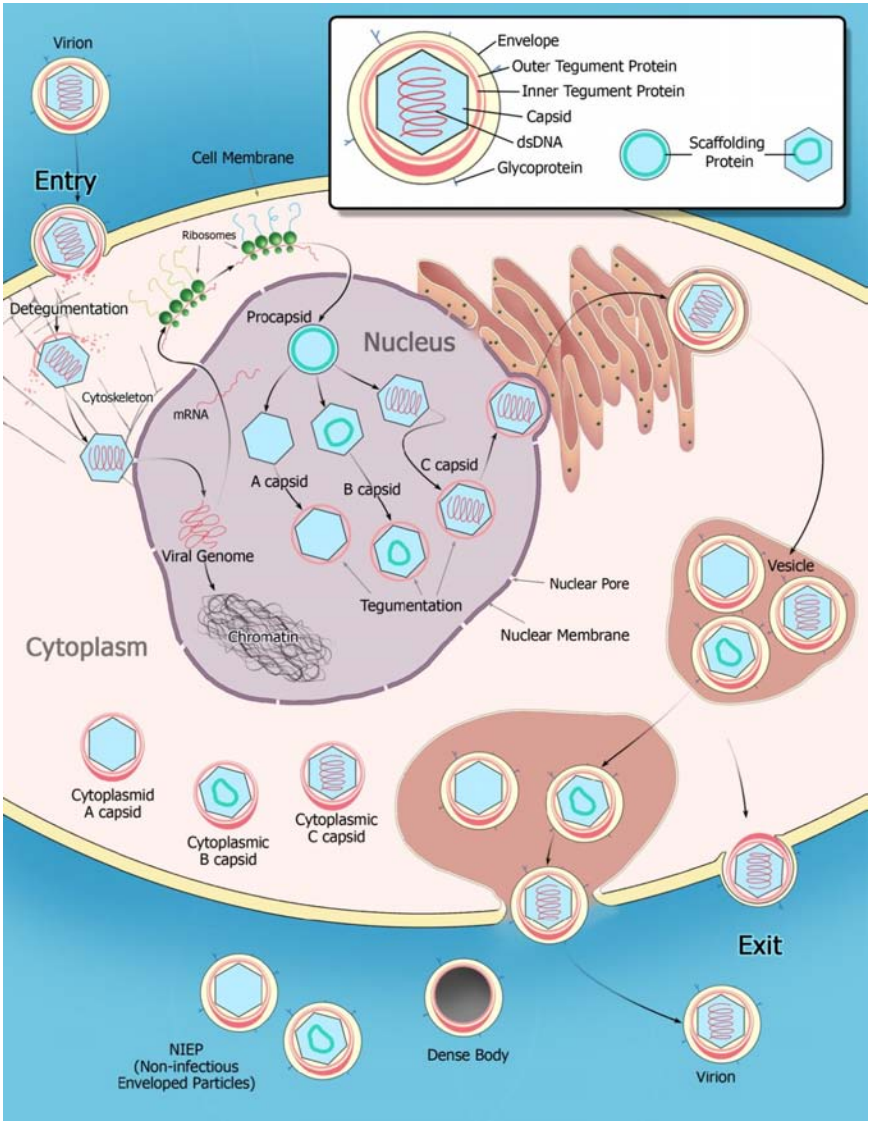


Fig. 1. Lytic replication cycle of herpesviruses. Upon recognition of the viral glycoprotein by a cell surface receptor, the viral envelope fuses with the host cell membrane, allowing the virus to enter the cell through either direct fusion or endocytosis. The virus is uncoated, releasing outer tegument proteins. The capsid, together with some inner tegument proteins, is transported

The procapsid can be assembled *in vitro* from capsid and scaffolding proteins in the absence of the viral protease.[24] Protease-minus procapsids can spontaneously rearrange into a large-cored, angular particle resembling the B-capsid, but these large-cored particles do not encapsidate DNA or become mature virions.[20,21,25] Cells infected with a temperature-sensitive, protease-minus HSV-1 mutant produce capsids that disassemble at the non-permissive temperature of 0°C, similar to the *in vitro*-assembled procapsids.[20] However, these capsids remain unstable at the permissive temperature and do not mature unless protease activity is restored.[20] The structures of the temperature-sensitive mutant, protease-minus mutant and *in vitro*-assembled capsids (Fig. 2) are all similarly porous and spherical compared to that of the wild-type A-capsid. These experiments illustrate the importance of protease cleavage in the herpesvirus capsid assembly and confirm that the procapsid is the precursor of the angular capsid. In fact, the interactions between the scaffolding protein (also known as assembly protein, or AP), major capsid protein (MCP), and viral protease have been explored as targets for antiviral drug design in herpesviruses.[26–30]

Further maturation of the capsids into infectious virions remains a controversial issue. One school of thought is that the capsids acquire their envelope from the inner nuclear membrane, process the envelope glycoproteins *in situ* and bud from the cell surface via the secretory pathway. However, new evidence suggests that HSV-1 capsids acquire their inner tegument proteins at or near the inner nuclear membrane and are enveloped temporarily, then fuse with the outer nuclear membrane to release tegumented capsids into the cytoplasm as illustrated in

---

to the nuclear pore to eject viral DNA into the nucleus, where DNA replication and RNA transcription occur. Viral proteins are translated in cytoplasm in stages and some are transported into the nucleus, where capsid assembly takes place. Spherical procapsids, which contain a ring of scaffolding proteins, are first assembled first. Spontaneous angularization coupled with DNA packaging lead to the formation of A, B, and C capsids. These capsids acquire their inner and outer tegument proteins at or near the nuclear membranes, enveloped, de-enveloped, and re-enveloped through the cellular excretion apparatus. Enveloped particles are released through exocytosis.

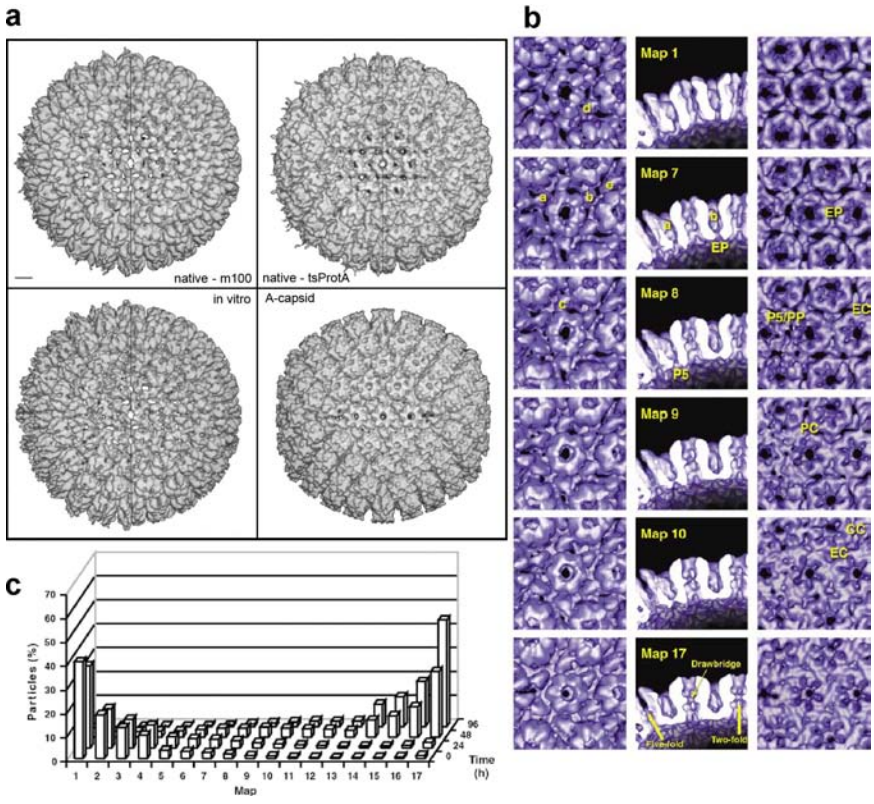


Fig. 2. Maturation of the HSV-1 capsid. (**a**) Protease-knockout mutant m100, temperature-sensitive mutant tsProtA, and *in vitro*-assembled procapsid all have a similar structure, which is more porous and spherical than that of the wild-type A capsid. (**b**) Capsid maturation is a continuous process with numerous intermediate steps (shown in chronological order from top to bottom). (**c**) Percentage of particles at various steps of maturation at given times. Adapted with permissions from the publishers and authors.[21,22]

Fig. 1. These tegumented capsids acquire their outer tegument layer proteins and envelope by budding into the *trans*-Golgi network.[31] In addition to electron micrographs showing capsids (presumably with a layer of inner tegument proteins) in the cytoplasm, Skepper *et al.* found that the phospholipid composition of secreted virions is different from that of virions in the perinuclear space and that the HSV-1

tegument protein VP22 is found exclusively in the cytoplasm of infected cells.[31] Viral nucleoid synthesis and envelope production may proceed independently under certain conditions, leading to overproduction and release of empty envelopes. While these particles are non-infectious, whether or not they can still interfere with cell function and/or autoimmunity remains to be studied.

Cryo-Electron Microscopy and Tomography as Tools for Herpesvirus Structure Determination

Herpesviruses are particularly challenging to structural biologists because of their large size, the structural complexity of their capsids, and the pleomorphic nature of their virion envelope and tegument layers. These unusual properties make the synergistic use of cryo-electron microscopy and single-particle reconstruction (commonly referred to collectively as "cryoEM") and cryo-electron tomography (cryoET) particularly important for structural studies of herpesviruses. Below, we will provide a general technical account of cryoEM and cryoET in herpesvirus structural studies, as well as specific examples of herpesvirus structures.

Although X-ray crystallography is the method of choice for revealing the atomic structures of large macromolecules and viruses, recent advances have given cryoEM an increasingly important role in determining subnanometer-resolution structures of such complexes.[32–36] Because the samples are frozen-hydrated rather than crystallized, cryoEM can reveal additional information about an entire virus' structure in its native form even when the crystal structures of viral components are already available.

Previously, cryoEM had been used only as a low-resolution technique (15–30 Å) due to the inevitable high level of noise in low-dose cryoEM images. In the last few years, however, much progress has made to demonstrate the feasibility of using cryoEM to resolve three-dimensional (3D) structures of icosahedral virus particles at subnanometer (6–9 Å) resolution.[32–35,37] This development represents a significant step forward, as *de novo* structure determination by cryoEM can now reveal new structural features such as α-helices and β-sheets[34–36,38] and uncover novel protein folds (see review by Zhou

and Chiu 2003).[35,36,39] If high-resolution crystallographic models of the individual proteins of a macromolecular complex are available (as is the case for the HSV-1 capsid structures), these atomic-resolution models can be fitted into the corresponding portions of the cryoEM map using the identified secondary structure elements as internal markers to guide the fitting, thereby establishing a pseudo-atomic model of the entire assembly. The large size of the herpesviruses (over 200 nm) and the presence of partially ordered tegument proteins inside the virion make cryoEM the method of choice for determining the structures of herpesvirus capsids and related particles. The use of higher-voltage cryo-electron microscopes — which offer greater penetrating power for reducing inelastically scattered electrons and increased depth of focus for imaging thicker specimens — is particularly important for imaging herpesvirus particles, which are among the largest animal virus studies to date. Improved data processing software, such as the IMIRS[40,41] and PFT packages[42] and others, also facilitates the automation and management of the large amount of data required to compute higher-resolution structures.

In recent years, cryoET has gained popularity as a tool for examining the 3D structures of macromolecular complexes that are too large, asymmetrical, or heterogeneous to be investigated by conventional cryoEM or crystallographic methods (reviewed by Subramaniam and Milne).[43] Rather than averaging the structures of many particles, as is the case in cryoEM, cryoET can be used to reconstruct a single, asymmetric particle without averaging.[44–48] Projection images of the ice-embedded sample are taken over a range of tilt angles. The 2D information from these images is merged to obtain 3D tomograms, which can be examined as a series of thin, serial slices. Specific molecules can be identified in cryoET reconstructions, indicating the potential for cryoET to approach true molecular resolution.[46,49] Recent advancements in high pressure freezing and cyro-sectioning[46,50,51] show great promise for studying intact, virus-infected cells using serial sections that are thin enough for tomographic studies. Enveloped viruses, such as herpesviruses, and their infection processes, can be visualized in 3D through tomographic reconstructions.

Structures of the Envelope and Tegument

All herpesviruses characterized to date share a characteristic multilayered architecture (Fig. 3a), composed of a dsDNA genome, a portal-containing icosahedral capsid shell, a thick, proteinaceous tegument compartment, and a lipid bilayer envelope spiked with glycoproteins.[52–54] CryoET reconstructions of purified HSV-1 virions by Grünewald *et al.* show pleomorphic or irregular oval shapes of the viral envelope with the capsid eccentrically located near one pole[47] (Fig. 3b). The HSV-1 viral membranes are smooth and contain about a dozen types of glycoprotein spikes, which appear to be more densely packed around the pole distal from the capsid.[47] This distribution suggests some sort of functional clustering that may help the virion engage cell receptors during infection.

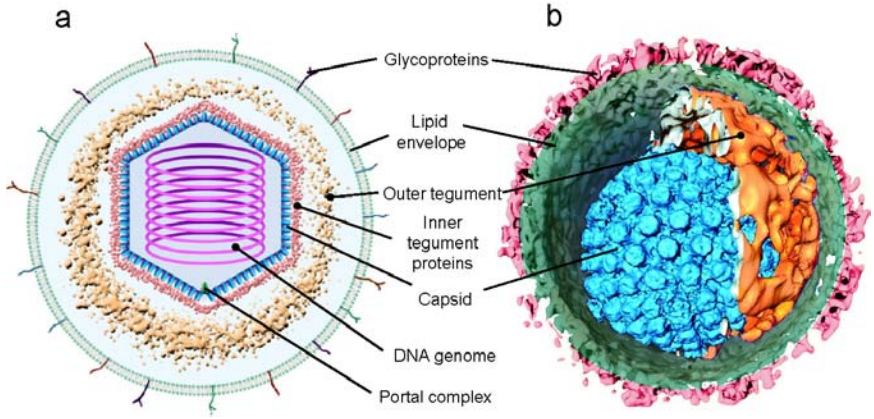


Fig. 3. Architecture of herpesvirus. (**a**) Basic architecture of the human herpesvirus virion. The dsDNA genome is coiled in the protein capsid. The capsid is surrounded by inner and outer tegument protein layers. The envelope is a host-derived lipid bilayer containing spikes of viral glycoproteins. (**b**) CryoET structure of HSV-1 virion. A montage of different structures is shown, including the envelope with glycoprotein (yellow), asymmetrical distribution of tegument densities (blue), and an averaged and icosahedrally symmetrized capsid. Panel b was adapted with permission from the publisher and the authors.[47]

The tegument is the compartment between the nucleocapsid and the envelope in herpesviruses and is thus architecturally equivalent to the matrix protein layer of other enveloped viruses, such as retroviruses and rhabdoviruses. Viral proteins in the tegument compartment are the least understood components, partly due to difficulties in isolating them. They play important functional roles in multiple aspects of the viral lifecycle, including capsid transport to the nucleus and virion assembly and egress.[53,55] The proximity of the tegument proteins to the surface of the viral capsid suggests that they may also be involved in modulating the interactions between viral capsids and the host cellular proteins upon entry into the cell.[56,57] In HSV-1, the tegument is asymmetrical and may contain actin-like filamentous substructures.[47]

Although there are several tegument proteins that are conserved in all herpesviruses, a large proportion of the tegument layer is apparently composed of proteins unique to each subfamily (Table 2).[58,59] The existence of distinct tegument layers among different subfamilies of herpesviruses (Fig. 4) has important biological implications in herpesvirus assembly and egress. Our structures of the betaherpesvirus tegument[60] and gammaherpesvirus tegument[61] clearly point to a divergence from that of HSV-1,[2] which is likely the result of divergence in the protein composition of the tegument. While inner tegument proteins in herpesviruses may be involved in essential processes in virion morphogenesis, such as capsid transport into and within the cytoplasm, outer tegument proteins are involved in modulating the host cell environment in the immediate-early phase of infection.[62,63] It is possible that during the course of herpesvirus evolution, some modulating proteins gained signals, allowing incorporation into the virion particle in a manner that did not disrupt the relatively conserved processes of capsid assembly, egress from the nucleus, and translocation to the site of virion assembly in the cytoplasm. The outer tegument layer is thus more divergent not only in structure, but also in its composition and function of its constituent proteins.

The organization and structure of the inner, or capsid-associated, tegument proteins varies greatly among herpesviruses across the herpesvirus subfamilies. Zhou *et al.* reconstructed to 20 Å the HSV-1

virion with an intact envelope and tegument (Fig. 4a).[2] The overall morphology of the virion is similar to that of the capsid, except for the presence of capsid-bound tegument densities. These appear as convoluted, ribbon-like structures that are anchored at one end to the pentons and extend to neighboring triplexes at the other end. The bulk of the attachment tegument density is centered on the pentons and makes only minimal contact with adjacent hexons.

In contrast, the tegument proteins of HCMV form a thin, net-like shell that surrounds the entire capsid (Fig. 4b).[60] The filamentous tegument densities contact the pentons, hexons and triplexes of the HCMV capsid. The structure of the human gammaherpesvirus tegument has not been reported because of difficulty in purifying intact virions. However, gammaherpesvirus tegument differs from those of HSV-1 and HCMV in that it does not make extensive symmetric contact with the capsid proteins.[61] Some portions of the inner tegument layer appeared to be tethered to the tips of the capsomers, but these contacts did not appear to be regular or to fill every possible binding site on the capsid. Also, in contrast to HSV-1, the majority of tegument molecules in virions of MHV-68, a mouse gammaherpesvirus, seem to be loosely organized, allowing the capsid to "float" inside the envelope and vary in its location within the tegument.[61]

Herpesvirus tegument proteins have essential functions in virion assembly and maturation through their interactions with the underlying capsid proteins and the overlying envelope glycoproteins. These proteins may also modulate the cellular machinery during the early stages of infection, as they are the first to contact the cellular environment after the viral envelope fuses with the cell membrane or the endocytic vesicle membrane and releases the partially tegumented nucleocapsid into the cytoplasm. The formation of tegument represents a subsequent evolutionary adaptation to infection of these nucleoplasms enclosed by a cytoplasm, when considering that herpesvirus capsids and icosahedral bacteriophages share a common ancestor that existed prior to the divergence of eubacteria and a primordial nucleoplasm.[64,65] With later divergence of the host range of herpesviruses, protein functions differentiated as well and have been incorporated into the tegument structures, as indicated by the structural

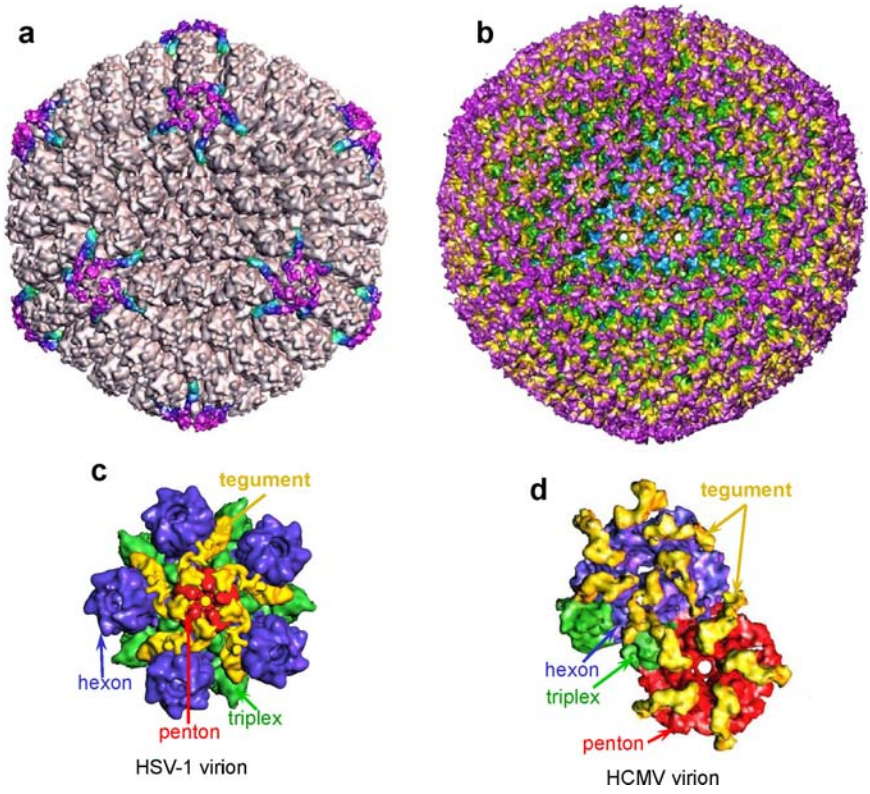


Fig. 4. Inner tegument proteins of HSV-1 and HCMV. (**a**) CryoEM reconstruction of HSV-1 virion showing the inner tegument proteins in color. (**b**) CryoEM reconstruction of HCMV virion showing inner tegument proteins in purple. (**c**) Close-up view of a computationally extracted portion of the HSV-1 virion reconstruction. The tegument densities, colored in yellow, make contact with the upper domains of the penton and its adjacent triplexes, but not the upper domains of any hexons. (**d**) Close-up view of a computationally isolated portion of the HCMV virion reconstruction. The tegument densities interact with the upper domains of all pentons, hexons, and triplexes by attaching one end to the upper domain of a hexon (or penton) and anchoring the other end on the upper domain of the adjacent triplex. Color codes in (**c**) and (**d**): red, penton; blue, hexon; green, triplex; yellow, tegument density.

results. Perhaps herpesviruses have achieved this evolutionary flexibility by varying the tegument protein composition and layering another, outer layer of tegument onto the inner layer associated with the capsid.

Structure of the Portal Complexes in Herpesvirus Capsids

Genetic and biochemical studies have suggested the existence of a bacteriophage-like, DNA-packaging/ejection portal complex in hepesviruses, presumably located at one of the 12 vertices of each capsid of herpesviruses.[66,67] These studies suggested that the HSV-1 portal exists at a single vertex of the capsid,[66] a finding that is analogous to the portal complex within DNA bacteriophages. Using cryoEM single-particle reconstruction, Trus and colleagues demonstrated that an isolated portal-like structure can self-assemble *in vitro* from purified UL6 protein of HSV-1.[67] However, attempts to visualize the putative herpesvirus portal complex in its native herpesvirus capsid environment by conventional single-particle cryoEM reconstruction approaches have all failed due to difficulties in identifying the unique vertex based on noisy cryoEM images by computational means.

Recently, three papers presented direct structural evidence for the existence of a portal complex in both HSV-1[68,69] and KSHV[70] capsids using the above discussed technique of cryoET (Fig. 5). Unfortunately, due to the inherent resolution limit and high noise level of current cryoET technology, the precise placement and orientation of the portal complex with respect to the capsid shell remains controversial and requires further study at an improved resolution. The internal placement of an "umbilicated" portal complex, similar to bacteriophage connectors, is supported by evidence from studies conducted by both Deng *et al.*[70] and Chang *et al.*,[69] which were based on structures of intact native KSHV capsids and chemically treated, penton-less HSV-1 capsids, respectively. In the latter case, high molar guanidine hydrochloride treatment was used to generate the penton-less HSV-1 capsids. Chang *et al.* demonstrated that the harsh chemical treatment did not

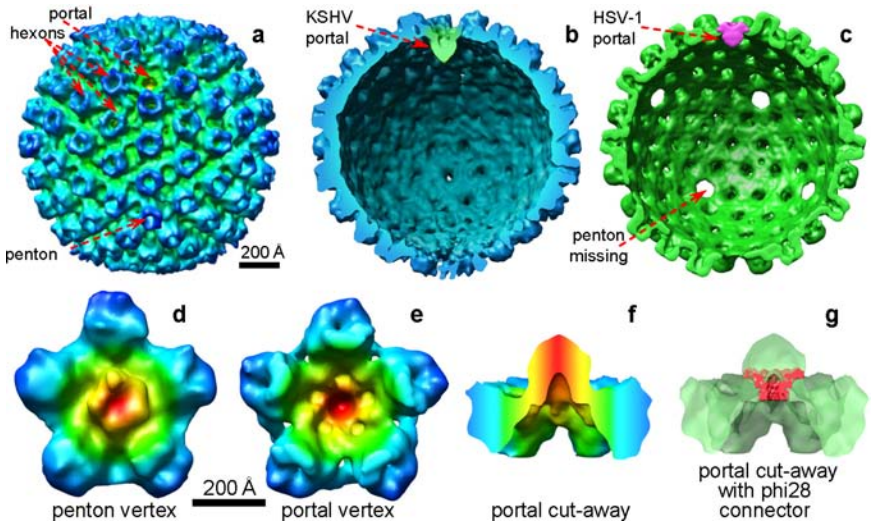


Fig. 5. Identification of portal complex in KSHV and HSV-1. (**a**) Radially colored surface representation of an averaged KSHV A-capsid cryoET reconstruction, as viewed along a two-fold axis. A unique portal vertex devoid of penton densities is identified at one of the 12 vertices. (**b**) Cut-away view of (**a**) showing the portal in green. (**c**) Cut-away view of the cryoET structure of chemically treated HSV-1 capsids.[69] Note: except for the portal vertex, all other vertices are unoccupied, indicating that all pentons were removed by the treatment. (**d**) One KSHV penton vertex with its five neighboring hexons. (**e**) The KSHV portal vertex and its five neighboring hexons. Maps in (**d**–**f**) are colored based on radial distance from the vertex axis. The densities closest to the vertex axis are shown in red, with the color changing first to green and then to blue to indicate densities that are further from the axis. (**f**) Side view of (**e**) with upper portion removed for clarity. (**g**) Fitting of the bacteriophage phi-29 portal connector complex at 10 Å (red) into the portal of KSHV portal vertex (semitransparent green). (Adapated with permision from publishers and the authors.[69–70])

remove the portal protein/complex along with the pentons. Notably, the KSHV capsids were not subjected to any such chemical extraction, so no ambiguities exist regarding the possible detachment of portal or portal-associated proteins.[70] The portal structure presented by Chang *et al.* in the chemically treated HSV-1 capsid is essentially identical to

the structure of the KSHV portal complex present in the KSHV native capsids (Fig. 5a).

The third related paper, by Cardone *et al.*, presented an averaged HSV-1 capsid map generated from alignment based on template matching over the tomograms.[68] It is possible that template matching introduced model bias, a problem made more severe in this particular case by the high level of noise present in cryoET tomograms. Moreover, the inherent missing-wedge problem due to limited tilt range in cryoET could have introduced structure distortion, which is quite likely to mislead template matching. This result is in agreement with the "docking experiments" in the paper published by the same group.[67] It is noteworthy that the docking was based on an assumption that the portal mass lies outside the capsid shell. This assumption stands contrary to the observations made in the native KSHV capsid[70] and the chemically treated HSV-1 capsid.[69]

Biochemical studies of KSHV particles have been particularly limited due to their low yield in culture and the consequent difficulty in viral particle isolation and purification. CryoET reconstructions of KSHV have provided the first direct evidence for the existence of a unique vertex or portal in KSHV or any gammaherpesviruses.[70] Indeed, the identification of the KSHV portal by cryoET reconstruction represents a new, alternative pathway of discovery using structural data, especially for proteins of lower abundance, when biochemical data remain scarce.

Capsid and Associated Structures of Human Herpesviruses

Among the many subviral particles and various enveloped particles present at various stages of the herpesvirus lytic replication cycle, the capsid is the most extensively studied structure to date. The herpesvirus capsid is particularly suitable for single-particle cryoEM reconstruction, thanks to the presence of icosahedral symmetry in the capsid shell (except for the presence of a unique portal vertex), the relative ease of isolation from the host nucleus, and the capsid shell's structural stability. Even at a moderate resolution of 20–30 Å, capsid subunits can be observed

undergoing conformational changes at different stages of capsid assembly.[22,71] The 3D folds of the protein subunits can be resolved at resolutions higher than 10 Å, allowing their interactions with DNA and other molecules to be revealed at the secondary structure level. Detailed comparisons between members of the herpesvirus family can provide clues as to how host specificity is determined and how viral assembly, maturation, and host cell interactions differ between herpesviruses. This and other structural information, such as the location of specific epitopes of viral proteins, may be useful in designing specific antiviral compounds.

Each herpesvirus capsid is composed of 12 pentons, 150 hexons, and 320 connecting triplexes, which are arranged on a T = 16 icosahedral lattice. The major capsid protein (MCP) forms the basic structural building block of the capsid. MCP proteins are arranged in groups of six ("hexons") or five ("pentons"), with the pentons forming the vertices of the icosahedron and the hexons filling up the faces. The hexons have slightly different geometries depending on their location, and are thus designated as P (penton-adjacent), E (edge), or C (center-face).[72] Connecting the pentons and hexons are triplexes, which act as molecular clamps and consist of one triplex monomer protein (TRI-1) and two triplex dimer proteins (TRI-2) each. There are six types of triplexes, designated Ta–Tf depending on their location in the asymmetric unit. Other capsid proteins include the small capsid protein (SCP), viral protease, and scaffold proteins (SCAF). The capsid shell of the herpesvirus capsid has a total mass of about 200 MDa, with 960 copies of MCP, 320 copies of TRI-1, 640 copies of TRI-2, and 900 copies of the SCP.

The capsid proteins of HSV-1, HCMV, and KSHV and their shared sequence homologies are summarized in Table 3. Below, we provide a detailed comparison of the capsid and capsid-associated tegument structures across HSV-1, HCMV, and KSHV, representatives of the alpha-, beta-, and gammaherpesvirus subfamilies, respectively.

HSV-1 Capsid

HSV-1 is the easiest of the human herpesviruses to culture and has been subjected to the most extensive structural analyses among all

Table 3. Capsid Proteins of HSV-1, HCMV, and KSHV, with Average Percent Identity as determined by CLUSTALW

Protein Function	Protein Name (Size in kDa)			Average % Identical Amino Acids
	HSV-1	HCMV	KSHV	
MCP	VP5 (149)	UL86 (154)	ORF25 (153)	26 ± 4
TDP	VP23 (34)	UL85 (35)	ORF26 (34)	16 ± 3
TMP	VP19c (50)	UL46 (33)	ORF62 (36)	15 ± 1
SCP	VP26 (12)	UL48.5 (8.5)	ORF65 (19)	6 ± 1

MCP, major capsid protein; TDP, triplex dimer protein; TMP, triplex monomer protein; SCP, smallest capsid protein.

herpesviruses. The highest-resolution capsid structure was obtained from over 5000 HSV-1 particle images recorded in a 400 kV microscope at 8.5 Å (Fig. 6a).[34] As the main building block of the HSV-1 capsid, the HSV-1 MCP, VP5, was the subject of two further studies. At 8.5-Å resolution, it is possible to visualize secondary structural elements that are not visible at lower resolutions. For example, α-helices appear as extended, cylindrical rods 6–8 Å in diameter (Fig. 6b). Bowman *et al.*[73] isolated a 65-kDa fragment from the upper domain of VP5 (VP5ud). They obtained a 2.9-Å crystal structure of the fragment, which they then fitted into the 8.5-Å cryoEM structure of the entire capsid. VP5ud was found to represent a novel protein fold not seen in any other virus families studied to date. The authors speculated that different conformations of the novel fold result in different electrostatic properties that allow VP5 subunits to form either hexons or pentons and may also mediate interactions of VP5 with other structural proteins. Subsequently, they identified 39 α-helices and four β-sheets from the VP5 subunits of the 8.5 Å cryoEM structure and validated them using the crystal structure of the VP5 upper domain.[74] The most striking observation was that 10 of the helices form a bundle in the middle domain of VP5 that resembles a domain fold found in the annexin protein family. This bundle is formed by the N- and C-terminal domains of VP5, according to structure-based sequence searches. Long helices in the floor domain of VP5 were also observed to form an interconnected network within and between capsomers and

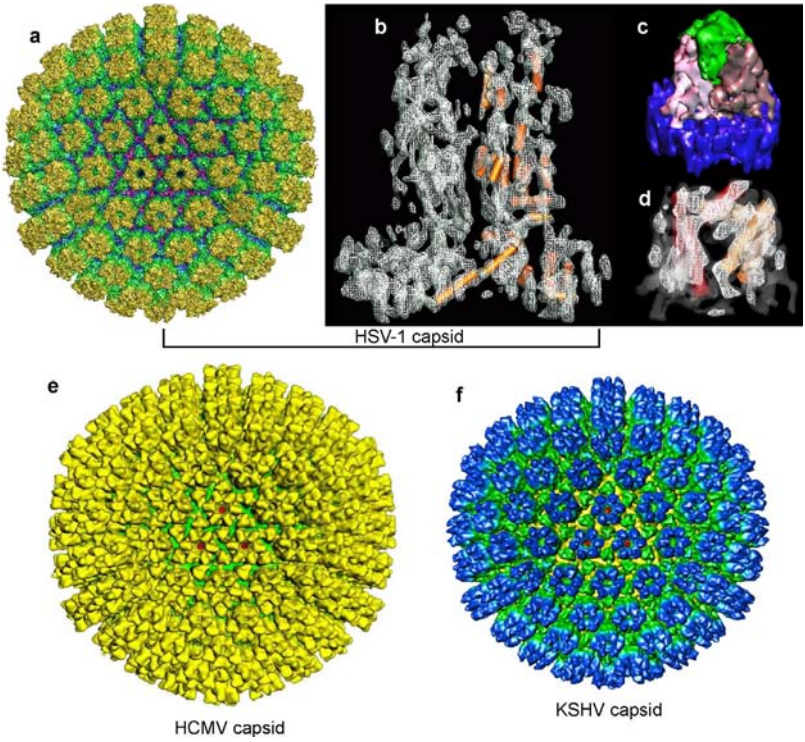


Fig. 6. Capsid structures of representative members of alpha-, beta-, and gammaherpesvirus. (**a**–**d**) Structural studies of HSV-1 capsid.[34] (**a**) 8.5 Å map of HSV-1 capsid. (**b**–**d**) Identification of helices in two neighboring MCP subunits (**b**) and in the triplex proteins (**c**–**d**). Helices are shown as 5 Å diameter cylinders. (**e**) CryoEM structure of HCMV capsid at 22 Å resolution.[23] (**f**) CryoEM structure of KSHV capsid at 24 Å resolution.[78] Adapted with permissions from the publishers.

resemble those seen in bacteriophage capsid protein.[64] This structural information was used to form an architectural model of the entire VP5 protein that can be fitted into the context of the whole capsid. Further structural studies of capsids at different stages of maturation will help elucidate the changes in helix arrangement during morphogenesis.

The 8.5 Å structure also gave deeper insight into the interesting quaternary structure of the triplexes (Figs. 6c and 6d), which are

composed of two molecules of VP23 and one molecule of VP19c. The lower portion of the triplex, which interacts with the floor of the pentons and hexons, is three-fold symmetric, with all three subunits roughly equivalent. This arrangement alters the middle of the triplex such that the upper portion is composed mostly of VP23 in a dimeric configuration. All three subunits of the triplex are required for the correct tertiary structure to form. VP23 exists in isolation only as a molten globule with no distinct tertiary structure.

HCMV Capsid

Structural studies of HCMV capsids have not reached as high a resolution as those of HSV-1, as it has been difficult to scale up and isolate an adequate amount of HCMV capsids from HCMV-infected cells. The first structure of the HCMV capsid, to 35 Å, was obtained by Butcher *et al.*[75] The hexons are slightly skewed with an oval, rather than round, axial channel. The penton tips are more pointed with a triangular cross-section than those of the hexons. Yu *et al.* subsequently obtained high-resolution structures that confirmed those features (Fig. 6e).[76,22]

The HCMV capsid structure is very similar to HSV-1 in overall organization, with four homologous structural proteins at the same stoichiometries. The main difference is that the HCMV capsid has a larger diameter (650 Å) than HSV-1 (620 Å), with a volume ratio of 1.17. The increased size of the HCMV capsid — despite the similar molecular mass of its component proteins — results in a greater center-to-center spacing of the capsomers and greater relative tilt between adjacent hexons compared to HSV-1.

The major point of interest in this study was the question of how HCMV packages its genome, which is 50% larger than that of HSV-1, into a capsid with only 17% more internal volume. Bhella *et al.* discovered that packing density is an important factor in genome encapsidation.[77] They compared radial density plots computed from cryoEM structures of HCMV and HSV-1. The average spacing between DNA layers in their HSV-1 reconstruction was 26 Å, whereas the interlayer

spacing for HCMV was 23 Å, indicating that the larger HCMV genome is more closely packed than that of HSV-1.

KSHV Capsid

Gammaherpesvirus structures are not as well studied as alpha- and betaherpesvirus structures due to difficulties in viral culturing and isolation of either EBV or KSHV. Wu *et al.* obtained the first capsid structure of a gammaherpesvirus, KSHV, in 2000 (Fig. 6f).[78] In these studies, KSHV in lymphoma cell cultures was induced to lytic replication using 12-*O*-tetradecanoylphorbol-13-acetate, yielding sufficient quantities of virus capsids for cryoEM studies. The 24 Å 3D structure of the KSHV capsid has the same T = 16 icosahedral symmetry as that of HSV-1 and HCMV capsids and has the same organization of 12 pentons, 150 hexons and 320 triplexes. The existence of KSHV A, B, and C capsids, similar to those of HSV-1, was further demonstrated subsequently.[79,80] These studies showed that the gross morphology of the capsid is highly conserved among herpesviruses, although the pentons, hexons, and triplexes differ slightly in shape.[80,81]

Despite success in pushing the resolution of the HSV-1 capsid beyond subnanometer resolution,[34] similar efforts for human beta- and gammaherpesviruses have met with difficulties. The purification process for HCMV and KSHV is still relatively low yield. Future assembly and structural studies of beta- and gammaherpesviruses toward higher resolution may depend on non-human viruses such as simian cytomegalovirus,[82] rhesus monkey rhadinovirus,[83] and murine gammaherpesvirus-68 (MHV-68).[59] These non-human pathogens replicate to high titer in culture, and their capsid, tegument, and envelope proteins share homology with those of HCMV (in the case of simian cytomegalovirus) and KSHV or EBV (in the case of rhesus monkey rhadinovirus and MHV-68). A recent study examined rhesus monkey rhadinovirus A, B, and C capsids in an attempt to understand gammaherpesvirus capsid maturation and DNA packaging.[83]

Location and Morphology of the SCP

As shown in Table 3, the most obvious difference among herpesvirus capsids is in the size of their small capsid protein (SCP). The SCPs are of great interest as they share the least sequence homology of herpesvirus capsid proteins.[78] The major and minor capsid proteins of herpesviruses share relatively high sequence similarity, so the overall capsid structure is very similar among the species, but the conformations, locations, and functions of the SCPs may be more diverse. The SCP of HCMV has been shown to be essential for HCMV infection *in vivo*,[84] whereas its counterpart in HSV-1, VP26, is dispensable for HSV-1 infection.[85,86] In KSHV, although the presence of SCP (pORF65) is used as a hallmark of KSHV infection for diagnostic purposes, it remains unknown whether it is essential for KSHV infection. However, HSV-1 VP26 associates with ribosomes and may regulate host protein translation.[87] A similar role for pORF65 is possible, as evidenced by another study in which pORF65 was shown to relocate from the nucleus to the cytoplasm of primary effusion lymphoma cells following induction of lytic replication.[88]

The SCPs have been difficult to study because of their small size and the relative low resolutions of the capsid KSHV and HCMV. At such resolutions, protein boundaries cannot be resolved, and it is difficult to recognize structural features of proteins as small as the SCP. In fact, even at 8.5 Å resolution, at which α-helices can be resolved, it has proven difficult to clearly delineate VP26 in HSV-1 hexons because of its extensive and interdigitated interactions across a large area of the MCP upper domain,[34] and because its secondary structure consists mainly of β-sheets.[89]

The 12 kDa SCP of HSV-1, VP26, binds the MCP subunits of the hexons, but not those of the pentons.[90] Its exact shape and location on the capsid — on the tips of the hexon subunits — have been determined by difference mapping[90] and observed in detail in the 8.5 Å structure.[34] However, a comparative study of wild-type and VP26-minus virions has shown that VP26 is not involved in the differential binding of tegument proteins to the pentons of the HSV-1 capsid.

It has not been possible to discern the location of KSHV SCP, pORF65, from the cryoEM structure,[78] although preliminary immunoEM experiments have confirmed that pORF65 is indeed bound to the capsid.[79] Instead, Lo *et al.* labeled KSHV capsids with anti-pORF65 antibodies and found that pORF65 binds only to the upper domains of the hexons and not those of the pentons,[93] similar to its counterpart in HSV-1.

Biochemical and immunoprecipitation studies by Lai and Britt have confirmed that the 8 kDa SCP of HCMV also interacts specifically with the MCP.[91] Based on architectural similarities, it was proposed that the HCMV SCP, like the SCPs of HSV-1 and KSHV,[90,92,93] binds only to MCP hexon subunits.[60,75] However, these suggestions could not be verified by structural studies because of the low resolutions of the HCMV capsid maps (~35 Å). Again, Yu *et al.* used anti-SCP antibodies to confirm that the SCP does indeed bind only to the hexons of HCMV.[94]

Given the very different sizes and amino acid sequences of the SCPs, it is not surprising that they should have different structural features. HSV-1 VP26 is a two-domain protein that forms interconnected, hexameric rings crowning the hexon.[90] The HCMV SCP attaches more toward the tips of the hexon subunits, giving them an elongated appearance.[75] The KSHV pORF65 appears to bind to the center of the hexon subunit upper domain, but it does not form the horn-shaped densities formed by VP26 on the HSV-1 hexons. These structural differences, along with the different roles of the SCPs in infection, illustrate how evolution can cause proteins to diverge in size, structure, and function even as their interactions and location in virus architecture are preserved.

Conclusion

Three-dimensional structural studies of human herpesviruses by cryoelectron microscopy and tomography have revealed both general similarities in their capsid assembly and structures, as well as differences in their tegument organization. These structural data have contributed to the understanding of their shared and unique characteristics with

regard to assembly, infection, replication, and pathogenesis. Future structural studies, particularly those at or near atomic resolution, will be essential to deciphering the mechanism of herpesvirus assembly and maturation, and will ultimately aid in developing therapeutic strategies for anti-herpes interventions.

Acknowledgments

We thank Ms. Xiaorui Zhang for graphical illustration and editorial assistance. The research described here was supported in part by grants from the National Institutes of Health (AI069015, CA094809, GM071940), the March of Dimes Birth Defects Foundation (5-FY99-852), the American Heart Association (0240216N), and the Welch Foundation.

References

1. Roizman B, Sears AE. (1996) Herpes simplex viruses and their replication. In B.N. Fields, D.M. Knipe, and P.M. Howley (eds), *Fields Virology*, **2**, pp. 2231–2295. Lippincott-Raven Publishers, Philadelphia.
2. Zhou ZH, Chen DH, Jakana J, *et al.* (1999) Visualization of tegument-capsid interactions and DNA in intact herpes simplex virus type 1 virions. *J Virol* **73**: 3210–3218.
3. Roizman B, Pellett PE. (2001) Herpesviridae: A Brief Introduction. In D.M. Knipe, P.M. Howley, D.E. Griffin, *et al.* (eds), *Fields Virology*, **2**, pp. 2381–2398. Lippincott Williams & Wilkins, Philadelphia.
4. Dalton AJ, Manaker RA. (1967) The comparison of virus particles associated with Burkitt's lymphoma with other herpes like viruses. In *Carcinogenesis: A Broad Critique*, pp. 59–90. Williams & Wilkins, Baltimore.
5. Biberfeld P, Kramarsky B, Salahuddin SZ, Gallo RC. (1987) Ultrastructural characterization of a new human B lymphotropic DNA virus (human herpesvirus 6) isolated from patients with lymphoproliferative disease. *J Natl Cancer Inst* **79**: 933–941.
6. Kramarsky B, Sander C. (1992) Electron microscopy of human herpesvirus-6 (HHV-6). In D.V. Ablashi, G.R.F. Krueger, S.Z. Salahuddin (eds), *Human Herpesvirus-6*, pp. 59–68. Elsevier Science Publications, Amsterdam.

7. Klussmann JP, Krueger E, Sloots T, *et al.* (1997) Ultrastructural study of human herpesvirus-7 replication in tissue culture. *Virchows Arch* **430**: 417–426.
8. Fingeroth JD, Clabby ML, Strominger JD. (1988) Characterization of a T-lymphocyte Epstein-Barr virus/C3d receptor (CD21). *J Virol* **62**: 1442–1447.
9. Lusso P, Secchiero P, Crowley RW, *et al.* (1994) CD4 is a critical component of the receptor for human herpesvirus 7: interference with human immunodeficiency virus. *Proc Natl Acad Sci USA* **91**: 3872–3876.
10. Santoro F, Kennedy PE, Locatelli G, *et al.* (1999) CD46 is a cellular receptor for human herpesvirus 6. *Cell* **99**: 817–827.
11. Greenstone HL, Santoro F, Lusso P, Berger EA. (2002) Human Herpesvirus 6 and Measles Virus Employ Distinct CD46 Domains for Receptor Function. *J Biol Chem* **277**: 39112–39118.
12. Cooper NR, Nowlin D, Taylor HP, Compton T. (1991) Cellular receptor for human cytomegalovirus. *Transplant Proc* **23**: 56–59, discussion 59.
13. Boyle KA, Compton T. (1998) Receptor-binding properties of a soluble form of human cytomegalovirus glycoprotein B. *J Virol* **72**: 1826–1833.
14. Olson JK, Grose C. (1998) Complex formation facilitates endocytosis of the varicella-zoster virus gE:gI Fc receptor. *J Virol* **72**: 1542–1551.
15. Akula SM, Pramod NP, Wang FZ, Chandran B. (2002) Integrin alpha3beta1 (CD 49c/29) is a cellular receptor for Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) entry into the target cells. *Cell* **108**: 407–419.
16. Dohner K, Wolfstein A, Prank U, *et al.* (2002) Function of dynein and dynactin in herpes simplex virus capsid transport. *Mol Biol Cell* **13**: 2795–2809.
17. Dalton AJ, Hageneau F. (1973) *Ultrastructure of Animal Viruses and Bacteriophages.* Academic Press, New York.
18. Heine UI. (1974) Intranuclear viruses. In (eds) *The Cell Nucleus*, **III**, pp. 489–536. Academic Press, New York.
19. Newcomb WW, Homa FL, Thomsen DR, *et al.* (1996) Assembly of the herpes simplex virus capsid: characterization of intermediates observed during cell-free capsid formation. *J Mol Biol* **263**: 432–446.
20. Rixon FJ, McNab D. (1999) Packaging-competent capsids of a herpes simplex virus temperature-sensitive mutant have properties similar to those of *in vitro*-assembled procapsids. *J Virol* **73**: 5714–5721.

21. Newcomb WW, Trus BL, Cheng N, *et al.* (2000) Isolation of herpes simplex virus procapsids from cells infected with a protease-deficient mutant virus. *J Virol* **74**: 1663–1673.
22. Heymann JB, Cheng N, Newcomb WW, *et al.* (2003) Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy. *Nat Struct Biol* **10**: 334–341.
23. Yu X, Trang P, Shah S, *et al.* (2005) Dissecting human cytomegalovirus gene function and capsid maturation by ribozyme targeting and electron cryomicroscopy. *Proc Natl Acad Sci USA* **102**: 7103–7108.
24. Newcomb WW, Homa FL, Thomsen DR, *et al.* (1999) Assembly of the herpes simplex virus procapsid from purified components and identification of small complexes containing the major capsid and scaffolding proteins. *J Virol* **73**: 4239–4250.
25. Zhou ZH, Macnab SJ, Jakana J, *et al.* (1998) Identification of the sites of interaction between the scaffold and outer shell in herpes simplex virus-1 capsids by difference electron imaging. *Proc Natl Acad Sci USA* **95**: 2778–2783.
26. Qiu X, Culp JS, DeLilla AG, *et al.* (1996) Unique fold and active site in cytomegalovirus protease. *Nature (London)* **383**: 275–279.
27. Shieh HS, Kurumbail RG, Stevens AM, *et al.* (1996) Three-dimensional structure of human cytomegalovirus protease. *Nature (London)* **383**: 279–282.
28. Tong L, Qian C, Massariol MJ, *et al.* (1996) A new serine-protease fold revealed by the crystal structure of human cytomegalovirus protease. *Nature* **383**: 272–275.
29. Flynn DL, Abood NA, Howerda BC. (1997) Recent advances in antiviral research: identification of inhibitors of the herpesvirus proteases. *Curr Opin Chem Biol* **1**: 190–196.
30. Tong L, Qian C, Massariol MJ, *et al.* (1998) Conserved mode of peptidomimetic inhibition and substrate recognition of human cytomegalovirus protease. *Nat Struct Biol* **5**: 819–826.
31. Skepper JN, Whiteley A, Browne H, Minson A. (2001) Herpes simplex virus nucleocapsids mature to progeny virions by an envelopment —> deenvelopment —> reenvelopment pathway. *J Virol* **75**: 5697–5702.
32. Böttcher B, Wynne SA, Crowther RA. (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* **386**: 88–91.

33. Conway JF, Cheng N, Zlotnick A, *et al.* (1997) Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature* **386**: 91–94.
34. Zhou ZH, Dougherty M, Jakana J, *et al.* (2000) Seeing the herpesvirus capsid at 8.5 Å. *Science* **288**: 877–880.
35. Zhou ZH, Baker ML, Jiang W, *et al.* (2001) Electron cryomicroscopy and bioinformatics suggest protein fold models for rice dwarf virus. *Nat Struct Biol* **8**: 868–873.
36. Zhou ZH, Zhang H, Jakana J, *et al.* (2003) Cytoplasmic polyhedrosis virus structure at 8 Å by electron cryomicroscopy: structural basis of capsid stability and mRNA processing regulation. *Structure* **11**: 651–663.
37. Trus BL, Roden RB, Greenstone HL, *et al.* (1997) Novel structural features of bovine papillomavirus capsid revealed by a three dimensional reconstruction to 9 Å resolution. *Nature Structural Biology* **4**: 411–418.
38. Jiang W, Li Z, Zhang Z, *et al.* (2003) Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nat Struct Biol* **10**: 131–135.
39. Zhou ZH, Chiu W. (2003) Determination of icosahedral virus structures by electron cryomicroscopy at subnanometer resolution. *Advances in Protein Chemistry* **64**: 93–124.
40. Liang Y, Ke EY, Zhou ZH. (2002) IMIRS: a high-resolution 3D reconstruction package integrated with a relational image database. *J Struct Biol* **137**: 292–304.
41. Dai W, Liang Y, Zhou ZH. (2003) Web portal to an image database for high-resolution three-dimensional reconstruction. *J Struct Biol* **144**: 238–245.
42. Baker TS, Cheng RH. (1996) A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *J Struct Biol* **116**: 120–130.
43. Subramaniam S, Milne JL. (2004) Three-dimensional electron microscopy at molecular resolution.* *Annu Rev Biophys Biomol Struct* **33**: 141–155.
44. Grimm R, Singh H, Rachel R, *et al.* (1998) Electron tomography of ice-embedded prokaryotic cells. *Biophys J* **74**: 1031–1042.
45. Nicastro D, Frangakis AS, Typke D, Baumeister W. (2000) Cryo-electron tomography of neurospora mitochondria. *J Struct Biol* **129**: 48–56.
46. Medalia O, Weber I, Frangakis AS, *et al.* (2002) Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science* **298**: 1209–1213.

47. Grunewald K, Desai P, Winkler DC, *et al.* (2003) Three-dimensional structure of herpes simplex virus from cryo-electron tomography. *Science* **302**: 1396–1398.
48. Grunewald K, Medalia O, Gross A, *et al.* (2003) Prospects of electron cryotomography to visualize macromolecular complexes inside cellular compartments: implications of crowding. *Biophys Chem* **100**: 577–591.
49. Sali A, Glaeser R, Earnest T, Baumeister W. (2003) From words to literature in structural proteomics. *Nature* **422**: 216–225.
50. Sartori Blanc N, Studer D, Ruhl K, Dubochet J. (1998) Electron beam-induced changes in vitreous sections of biological samples. *J Microsc* **192** (Pt 2): 194–201.
51. Luther PK, Morris EP. (2003) Cryoelectron microscopy of refrozen cryosections. *J Struct Biol* **142**: 233–240.
52. Rixon FJ. (1993) Structure and assembly of herpesviruses. *Seminars in Virology* **4**: 135–144.
53. Steven AC, Spear PG. (1997) Herpesvirus capsid assembly and envelopment. In W. Chiu, R.M. Burnett, R. Garcea (eds), *Structural Biology of Viruses*, pp. 312–351. Oxford University Press, New York.
54. Liu F, Zhou ZH. (2005) Comparative virion structures of human herpesviruses. In A. Arvin, G. Campadelli-Fiume, P. Moore, *et al.* (eds), *Human Herpesviruses: Biology, Therapy and Immunoprophylaxis*, in press. Cambridge University Press, Cambridge, UK.
55. Mettenleiter TC. (2002) Herpesvirus assembly and egress. *J Virol* **76**: 1537–1547.
56. Morrison EE, Stevenson AJ, Wang YF, Meredith DM. (1998) Differences in the intracellular localization and fate of herpes simplex virus tegument proteins early in the infection of Vero cells. *J Gen Virol* **79**: 2517–2528.
57. Castillo JP, Kowalik TF. (2002) Human cytomegalovirus immediate early proteins and cell growth control. *Gene* **290**: 19–34.
58. Kieff E, Rickinson AB. (2001) Epstein-Barr virus and its replication. In D.M. Knipe, P.M. Howley, D.E. Griffin, *et al.* (eds), *Fields' Virology*, **2**, pp. 2511–2574. Lippincott Williams & Wilkins, Philadephia, USA.
59. Bortz E, Whitelegge JP, Jia Q, *et al.* (2003) Identification of proteins associated with murine gammaherpesvirus 68 virions. *J Virol* **77**: 13425–13432.
60. Chen DH, Jiang H, Lee M, *et al.* (1999) Three-dimensional visualization of tegument/capsid interactions in the intact human cytomegalovirus. *Virology* **260**: 10–16.

61. Dai W, Jia Q, Bortz E, *et al.* (2007) Cryo-electron microscopy and tomography reveal distinct tegument-capsid interactions in a tumor herpesvirus. *J Struct Biol* In press.
62. Zhu FX, Yuan Y. (2003) The ORF45 protein of Kaposi's sarcoma-associated herpesvirus is associated with purified virions. *J Virol* **77**: 4221–4230.
63. Jia Q, Chernishof V, Bortz E, *et al.* (2005) Murine gammaherpesvirus 68 open reading frame 45 plays an essential role during the immediate-early phase of viral replication. *J Virol* **79**: 5129–5141.
64. Baker ML, Jiang W, Rixon FJ, Chiu W. (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. *J Virol* **79**: 14967–14970.
65. Jiang W, Chang J, Jakana J, *et al.* (2006) Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* **439**: 612–616.
66. Newcomb WW, Juhas RM, Thomsen DR, *et al.* (2001) The UL6 gene product forms the portal for entry of DNA into the herpes simplex virus capsid. *J Virol* **75**: 10923–10932.
67. Trus BL, Cheng N, Newcomb WW, *et al.* (2004) Structure and polymorphism of the UL6 portal protein of herpes simplex virus type 1. *J Virol* **78**: 12668–12671.
68. Cardone G, Winkler DC, Trus BL, *et al.* (2007) Visualization of the herpes simplex virus portal *in situ* by cryo-electron tomography. *Virology* **361**: 426–434.
69. Chang JT, Schmid MF, Rixon FJ, Chiu W. (2007) Electron cryotomography reveals the portal in the herpesvirus capsid. *J Virol* **81**: 2065–2068.
70. Deng B, O'Connor CM, Kedes DH, Zhou ZH. (2007) Direct visualization of the putative portal in the Kaposi's sarcoma-associated herpesvirus capsid by cryoelectron tomography. *J Virol* **81**: 3640–3644.
71. Trus BL, Booy FP, Newcomb WW, *et al.* (1996) The herpes simplex virus procapsid: structure, conformational changes upon maturation, and roles of the triplex proteins VP19c and VP23 in assembly. *J Mol Biol* **263**: 447–462.
72. Steven AC, Roberts CR, Hay J, *et al.* (1986) Hexavalent capsomers of herpes simplex virus type 2: symmetry, shape, dimensions, and oligomeric status. *J Virol* **57**: 578–584.
73. Bowman BR, Baker ML, Rixon FJ, *et al.* (2003) Structure of the herpesvirus major capsid protein. *EMBO J* **22**: 757–765.

74. Baker ML, Jiang W, Bowman BR, *et al.* (2003) Architecture of the herpes simplex virus major capsid protein derived from structural bioinformatics. *J Mol Biol* **331**: 447–456.
75. Butcher SJ, Aitken J, Mitchell J, *et al.* (1998) Structure of the human cytomegalovirus B capsid by electron cryomicroscopy and image reconstruction. *J Struct Biol* **124**: 70–76.
76. Yu X, Shah S, Atanasov I, *et al.* (2005) Three-dimensional localization of the smallest capsid protein in the human cytomegalovirus capsid. *J Virol* **79**: 1327–1332.
77. Bhella D, Rixon FJ, Dargan DJ. (2000) Cryomicroscopy of human cytomegalovirus virions reveals more densely packed genomic DNA than in herpes simplex virus type 1. *J Mol Biol* **295**: 155–161.
78. Wu L, Lo P, Yu X, *et al.* (2000) Three-dimensional structure of the human herpesvirus 8 capsid. *J Virol* **74**: 9646–9654.
79. Nealon K, Newcomb WW, Pray TR, *et al.* (2001) Lytic replication of Kaposi's sarcoma-associated herpesvirus results in the formation of multiple capsid species: isolation and molecular characterization of A, B, and C capsids from a gammaherpesvirus. *J Virol* **75**: 2866–2878.
80. Trus BL, Heymann JB, Nealon K, *et al.* (2001) Capsid structure of Kaposi's sarcoma-associated herpesvirus, a gammaherpesvirus, compared to those of an alphaherpesvirus, herpes simplex virus type 1, and a betaherpesvirus, cytomegalovirus. *J Virol* **75**: 2879–2890.
81. Wu L, Stoops JS, Forghani B, Zhou ZH. (1999) *Three-dimensional structure of the human herpesvirus 8 capsid*. 24th International Herpesvirus Workshop, MIT, Boston, MA.
82. Trus BL, Gibson W, Cheng N, Steven AC. (1999) Capsid structure of simian cytomegalovirus from cryoelectron microscopy: evidence for tegument attachment sites. *J Virol* **73**: 2181–2192.
83. Yu XK, O'Connor CM, Atanasov I, *et al.* (2003) Three-dimensional structures of the A, B, and C capsids of rhesus monkey rhadinovirus: insights into gammaherpesvirus capsid assembly, maturation, and DNA packaging. *J Virol* 77: 13182–13193.
84. Borst EM, Mathys S, Wagner M, *et al.* (2001) Genetic evidence of an essential role for cytomegalovirus small capsid protein in viral growth. *J Virol* **75**: 1450–1458.
85. Desai P, DeLuca NA, Person S. (1998) Herpes simplex virus type 1 VP26 is not essential for replication in cell culture but influences pro-

duction of infectious virus in the nervous system of infected mice. *Virology* **247**: 115–124.
86. Chen DH, Jakana J, McNab D, *et al.* (2001) The pattern of tegument-capsid interaction in the herpes simplex virus type 1 virion is not influenced by the small hexon-associated protein VP26. *J Virol* **75**: 11863–11867.
87. Greco A, Bienvenut W, Sanchez JC, *et al.* (2001) Identification of ribosome-associated viral and cellular basic proteins during the course of infection with herpes simplex virus type 1. *Proteomics* **1**: 545–549.
88. Katano H, Sato Y, Kurata T, *et al.* (2000) Expression and localization of human herpesvirus 8-encoded proteins in primary effusion lymphoma, Kaposi's sarcoma, and multicentric Castleman's disease. *Virology* **269**: 335–344.
89. Wingfield PT, Stahl SJ, Thomsen DR, *et al.* (1997) Hexon-only binding of VP26 reflects differences between the hexon and penton conformations of VP5, the major capsid protein of herpes simplex virus. *Journal of Virology* **71**: 8955–8961.
90. Zhou ZH, He J, Jakana J, *et al.* (1995) Assembly of VP26 in herpes simplex virus-1 inferred from structures of wild-type and recombinant capsids. *Nat Struct Biol* **2**: 1026–1030.
91. Lai L, Britt WJ. (2003) The interaction between the major capsid protein and the smallest capsid protein of human cytomegalovirus is dependent on two linear sequences in the smallest capsid protein. *J Virol* **77**: 2730–2735.
92. Trus BL, Homa FL, Booy FP, *et al.* (1995) Herpes simplex virus capsids assembled in insect cells infected with recombinant baculoviruses: structural authenticity and localization of VP26. *J Virol* **69**: 7362–7366.
93. Lo P, Yu X, Atanasov I, *et al.* (2003) Three-Dimensional Localization of pORF65 in Kaposi's Sarcoma-Associated Herpesvirus Capsid. *J Virol* **77**: 4291–4297.
94. Yu X, Shah S, Atanaso I, *et al.* (2005) Three-dimensional localization of the smallest capsid protein in the human cytomegalovirus capsid. *J Virol* **79**: 1327–1332.

Chapter 19

Human Papillomavirus Type 16 Capsid Proteins: Immunogenicity and Possible Use as Prophylactic Vaccine Antigens

Tadahito Kanda[*,†], *Kei Kawana*[‡] *and Hiroyuki Yoshikawa*[§]

Human papillomavirus 16 (HPV16), a small non-enveloped DNA virus, is one of the HPV types that causes cervical cancer, accounting for 50% of the cases. Although cell cultures allowing HPV16 propagation are not available, surrogate systems can provide the HPV16 capsids and infectious pseudovirions, which have facilitated the biological and immunological studies of capsid proteins. HPV16 major capsid protein L1 and minor capsid protein L2 are supposed to be involved in the attachment to the cell surface and the intracellular transport, respectively, in the viral infection. L1 and L2 seem to elicit type-specific and type-common, respectively, neutralizing antibodies. A recent clinical trial in the United States[1] strongly suggests that HPV16 L1-capsids can serve as a prophylactic vaccine against HPV16 infections, some of which would progress to malignancy. Studies are underway on the possible use of L2 for a vaccine against multiple mucosal HPVs.

*Corresponding author.
†Division of Molecular Genetics, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan, E-mail: kanda@nih.go.in
‡Department of Obstetrics and Gynecology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan.
§Department of Obstetrics and Gynecology, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba-shi, Ibaraki 305-8575, Japan.

Introduction

Since its first discovery in cottontail rabbits in the early 1930s, papillomavirus, a small non-enveloped DNA virus, has been isolated from the human, monkey, bovine, rabbit, deer, hamster, and so on. These papillomaviruses have a highly species-specific host range and thus are named together with the host animal name, e.g. cottontail rabbit papillomavirus (CRPV). The viruses attracted a lot of attention because they cause a variety of proliferating epithelial lesions, such as benign papilloma, wart, condyloma, and carcinoma. However, detailed biological studies of the viruses were greatly hampered, largely due to the lack of a cell culture system for virus propagation, until the viral genomes were molecularly cloned in the late 1970s. The cloning allowed the standardization of viral reagents and provided sufficient material to initiate systematic studies of these viruses.[2]

Among them, bovine papillomavirus type 1 (BPV1), which causes fibrosarcoma in cattle, had been studied most extensively, largely owing to the introduction of an infectivity assay system (focus formation of transformed cells) using rodent cells. The molecular studies provided us with a general understanding of the papillomavirus genomic structure and gene functions.[2]

The molecular cloning of human papillomavirus in the 1980s (HPV6 and HPV11 from benign genital warts; HPV16 and HPV18 from cervical carcinoma biopsies) facilitated the biological and epidemiological studies and a systematic evaluation of various HPVs.[52,53] To date more than 100 genotypes (classified by DNA homology) have been cloned and are grouped into mucosal and cutaneous types based on the tissue tropism. Among the mucosal types, HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 and 73 are believed to be associated with cervical cancer, the second most frequent gynecological malignancy in the world, and HPV16 accounts for 50% of the cases.

Morphological and chemical studies were conducted with BPV1 and HPV1 virions isolated from the clinical lesions.[2,4] The virion (with a diameter of 55 nm) comprises an icosahedral non-enveloped protein shell designated as capsid and a single copy of the 8 kb double-stranded circular DNA genome contained within the capsid. The capsid consists of major capsid protein L1 (mol. wt.: 55,000) and minor capsid

protein L2 (mol. wt.: 76,000). Seventy-two capsomeres (each is an L1-pentamer) are arranged on a skew T = 7d icosahedral surface lattice of the capsid. Twelve capsomeres contact five neighbors and the others, six. Three-dimensional image reconstruction suggests that L2 is located at the center of a pentavalent capsomere at the virion vertices.

Histological studies of the epithelium infected with HPVs showed that the HPV genome is maintained as an episome in the nuclei of basal cells, where the viral gene expression is largely suppressed.[3] When the infected cells initiate terminal differentiation, the HPV genome starts to be transcribed and replicates.[3] In the upper layers of the epidermis or mucosa, HPV virions are generated and released.[3] Thus, the HPV lifecycle is closely associated with the epithelial differentiation, and likely accounts for fact that there is no cell culture system for HPV virion production.

Despite the lack of cell culture systems for HPV propagation, surrogate systems have been developed to produce HPV virion proteins; capsid proteins L1 and L2; the capsids comprise L1 alone (L1-capsids, also called L1 virus-like particle or L1-VLPs); the capsids comprised of L1 and L2 (L1/L2-capsids, also called L2-VLPs); and infectious pseudovirions that contain reporter DNA. These materials have permitted us to study the biological functions and immunogenicity of the capsid proteins. Given the role of HPV16 in carcinogenesis, we have focused on the recent studies of HPV16 capsid proteins in this review.

HPV Capsids and Pseudovirions

Surrogate Production of HPV Capsids

Zhou *et al.*[5] introduced a substituting method to produce HPV16 empty virus-like particles in cell cultures. HPV16 L1, when expressed alone or together with L2 from a recombinant vaccinia virus vector, self-assembles into L1-capsids or L1/L2-capsids, respectively, in the nuclei of infected monkey CV-1 cells. Hagensee *et al.*[6] have demonstrated the formation of HPV1 L1- and L1/L2-capsids by a similar approach.

Kirnbauer *et al.*[7,8] showed that self-assembled BPV1 and HPV16 L1-capsids can be obtained from the expression of recombinant bac-

uloviruses in insect sf9 cells. Subsequently, HPV16, HPV11, and HPV33 were shown to yield self-assembled L1/L2-capsids in the similar way.[9,10] These particles can be easily purified by a CsCl-equilibrium density-gradient centrifugation. They are morphologically indistinguishable, under an electron microscope, from the authentic HPV virions extracted from the lesions.[6,11]

Recently, Leder *et al.*[12] introduced another method to produce HPV16 L1/L2-capsids, using an expression plasmid with codon-modified L1 and L2 genes. The modification, which entails certain nucleotide substitutions without changing the amino acid sequences in the two proteins, greatly enhanced the expression levels of L1 and L2 in human 293T cells transfected with the expression plasmid. Similarly, L1 and L2 of HPV6, 18, 31, and 52 were found to be efficiently expressed and to yield L1/L2-capsids in the cells transfected with expression plasmids having the modified L1 and L2 genes (our unpublished data).

It turned out that HPV L1 and L2 mRNAs transcribed from the unmodified genes are extremely unstable in mammalian cells transfected with expression plasmids. Collier *et al.*[13] and Oeben *et al.*[14] mapped the cis elements that cause the mRNA instability to the 5′ 500-nucleotide regions of L1 and L2 open reading frames. It is likely that the L1 and L2 mRNAs transcribed by RNA polymerase II in the nuclei are degraded by an unidentified cellular function, which presumably accounts for the lack of HPV capsid production in the cells transfected with the authentic capsid genes. In the vaccinia vector system, L1 and L2 mRNAs are not degraded, probably because they are transcribed by the viral RNA polymerase in the cytoplasm, and insect cells lack the putative degradation function. The instability of the mRNAs and the resulting suppression of capsid protein expression appear to be related to the lifecycle of HPV, because L1 and L2 are expressed from the authentic genes in differentiating keratinocytes (our unpublished data).

Disulfide Bonds Required for Capsid Assembly

McCarthy *et al.*[15] showed that reducing agents cause disassembly of HPV11 L1-capsids to the level of the capsomeres, indicating that the capsid assembly is dependent on intercapsomeric disulfide bonds.

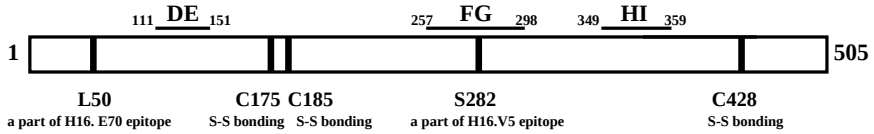


Fig. 1. Amino acid residues associated with L1 functions. Numbers represent amino acid numbers of HPV16 L1. DE, FG, and HI are HPV16 capsid surface loops.[17] C175 (cysteine at aa175), C185, and C428 are required for the normal assembly of L1-capsids.[16] L50 and S282 are located in conformationally dependent epitopes for MAbs H16.E70[39] and H16.V5.[41]

We have studied a series of HPV16 L1 mutants having nucleotide substitution of serine for cysteine and identified cysteine residues required for the intercapsomeric disulfide bonds[16] (Figs. 1 and 2). HPV16 L1 comprises 505 amino acids (aa), including 12 cysteines. The nuclear extracts from sf9 cells expressing the HPV16 L1 mutants have been examined for the presence of intermolecular bonding by SDS-PAGE with or without a reducing agent. Wild-type (WT) capsid is separated into two bands of L1-trimers and L1-dimers in the non-reducing SDS gel. C175S (substitution for cysteine at aa175) and C185S do not form the dimer and C428S does not form the trimer or the dimer.

Figure 2 shows structures formed by the mutants (photos taken under a transmission electron microscope). WT forms typical L1-capsids with a diameter of 55 nm. C175S forms mainly heterogeneous rod-shaped tubular structures. C185S forms capsid-like particles mostly smaller than WT, with diameters of 40 to 55 nm. C428S contains only capsomeres, indicating that lack of trimerization and dimerization does not affect capsomere formation. The three cysteines are thought to be localized in the loops displayed on the surface by the X-ray crystallographic analysis of the capsid-like structures composed of truncated L1 of HPV16.[17,18]

Although it is not clear which pairs of cysteines would make the intercapsomeric disulfide bonds for correct assembly of capsids, C428 may participate in intercapsomeric bonding of all capsomeres, and C175 and C185 may be involved in bonding the complex of three and two capsomeres, respectively. Two types of junctions of capsomeres

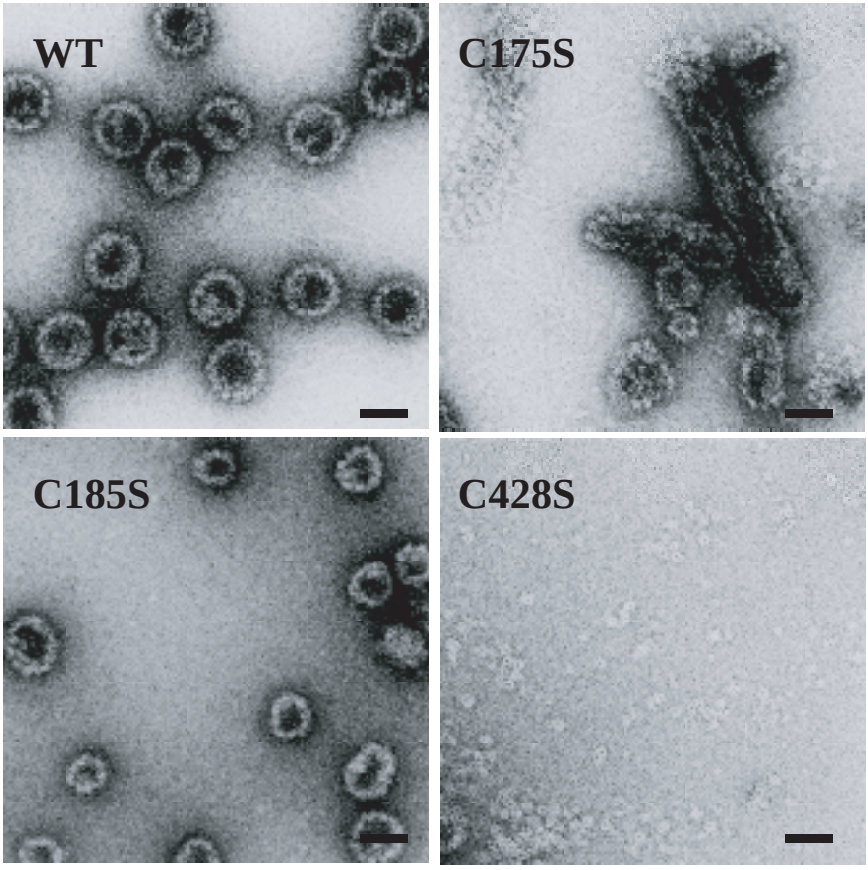


Fig. 2. Electron micrographs of particles formed by HPV16 L1 (WT), C175S (the HPV16 L1 mutant with substitution of S for C at amino acid 175), C185S, and C428S.[16] The particles are obtained by expression of the L1 genes with or without mutations in insect sf9 cells from recombinant baculoviruses.

may be necessary to form an icosahedral structure composed of pentavalent and hexavalent capsomeres.

The locations of cysteines are well-conserved in the L1 proteins from various HPVs. The importance of C175 and C428 in HPV16 L1 is consistent with the results of the studies of cysteine mutants of HPV33.[19]

L2 in HPV Capsids

Although the papillomavirus capsid can be assembled from L1 alone, the virion always contains L2 as a minor component in the center of a pentavalent capsomere at the virion vertices.[18] It is as yet to be investigated how L2 and L1 interact with each other in the formation of the particles.

Zhou *et al.*[20] showed that the amino-terminus of HPV16 L2, which is located inside the capsid, is able to bind DNA, suggesting a role for L2 in attracting viral DNA during the assembly of viral particles. So far the specific nucleotide sequence for efficient binding with L2 has not been found.

We showed that mouse monoclonal antibodies recognizing epitopes within regions of amino acid (aa) 69–81 and aa108–120 of HPV16 L2 bind to the L1/L2-capsids, indicating that the regions are located on the surface of the capsid.[21]

Production of Infectious Pseudovirions

Several methods have been developed to prepare infectious HPV pseudovirions by encapsidating reporter plasmid DNA into L1/L2-capsids. The pseudovirions are used in place of HPV virions to study the functions of the capsid proteins in viral infection and the neutralizing activities of anti-HPV antibodies.

Roden *et al.*[22] produced HPV16 pseudovirions by packaging BPV1 DNA in BPHE-1 cells — a hamster fibroblast line harboring the BPV1 genomic DNA as autonomously replicating episomes, but not expressing the BPV1 capsid protein genes — by using defective Semliki Forest virus as an expression vector. Co-expression of HPV16 L1 and L2 in the cells results in production of HPV16 L1/L2-capsids containing BPV1 genomic DNA. The infectivity can be assayed by counting the number of foci of transformed cells in the mouse C127 culture infected with the pseudovirions.

Recently, Buck *et al.*[23] introduced a system to produce pseudovirions easily by using the plasmids encoding the codon-modified HPV16 L1 and L2 genes and the cells expressing SV40 T antigen, such as

COS-1 and 293T. The transfection of the cells with the L1 and L2 expression plasmids together with a reporter plasmid with the SV40 replication origin (pSV plasmid) produces the L1/L2-capsids containing reporter DNA replicating in the nuclei of the cells. When secreting alkaline phosphatase is used as a reporter, the resulting pseudovirions provide a highly sensitive neutralization assay method for serum antibodies against HPV16 L1.

In the above-mentioned pseudovirion production, L2 is required for efficient encapsidation of DNA.[22,23] The other viral nonstructural proteins including E2 are not required to enhance the production of the pseudovirions.[23] The specific DNA sequences required for packaging, such as packaging signals, have not yet been found.

We developed a method for encapsidating plasmid DNA into HPV16 L1/L2-capsids in a cell-free system.[24] The L1/L2-capsids that have been disassembled to the level of capsomeres with reducing agent 2-mercaptoethanol (2-ME) (5%) are mixed with purified pSV2 plasmid DNA expressing a reporter gene, and then 2-ME is removed by dialysis to allow the capsomeres to reassemble into L1/L2-capsids. Some of the reassembled capsids package the plasmid to form pseudovirions. The pseudovirions produced by reassembling in the cell-free system are morphologically similar to the original L1/L2-capsids.

Infection with the pseudovirions can be monitored by the expression of the reporter gene. The pseudovirions provide tools to examine antibodies for their capability to inhibit HPV infection and to study early events in the infection, such as cell binding and internalization. However, when cells are infected with the pseudovirions at a low multiplicity of infection (MOI), the expression is usually difficult to detect. The autonomous replication of the pSV2 reporter plasmid mediated by SV40 T-antigen in particular cells, such as COS-1 and 293T, is required for efficient detection of the infected cells.

Interaction of HPV Capsids with Cell Surface

Cellular Attachment Receptor for HPV

Viral infection is supposed to start from the attachment of virions to the cell surface. The adsorption of HPV appears to occur from the

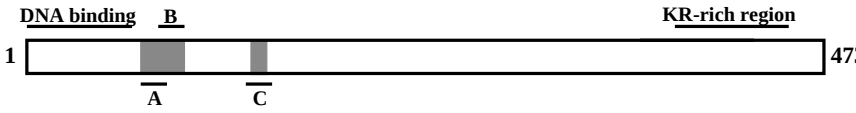
binding of L1 to cell receptors without the involvement of L2, because virions of BPV1,[25] as well as L1/L2- and L1-capsids of BPV1, HPV11, HPV16, and HPV33,[25–27] are seemingly capable of binding to cells with a similar efficiency, and because a mouse anti-BPV1 L2 monoclonal antibody, which inhibits the infectivity of BPV1 (focus formation in mouse C127 cells), allows the binding of BPV1 virions to C127 cells.[28]

Giroglou *et al.*[29] and Shafti-Karamat *et al.*[30] showed that a cellular receptor for HPV L1 is cell surface heparan sulfates. L1-capsids of several HPV types bind to cell surface heparan sulfates and heparin, which is a highly sulfated form of heparan sulfate proteoglycan. Heparin interferes with the infection of COS-7 cells with HPV16 and HPV33 pseudovirions and the infection of human primary keratinocytes with authentic HPV11 virions. COS-7 cells whose surface heparan sulfates have been removed by heparinase I treatment are not infected with HPV16 and 33 pseudovirions. Erythroleukemia cell line K562, which does not express syndecans at the cell surface, is not infected with HPV11 virions, and syndecan-1-positive K562 transfectants produced by an expression plasmid for syndecan-1 are sensitive to HPV11 infection. The results clearly indicate that heparan sulfate proteoglycans are a cellular attachment receptor for HPV.

L2 Functions in the Initial Steps of HPV Infection

Yang *et al.*[31] reported the data suggesting that L2 plays essential roles in the intracellular transport of virions to the nucleus. They compared the binding, uptake, and intracellular transport of HPV16 L1/L2-capsids with those of L1-capsids by transmission electron microscopy. HPV16 L1-capsids and L1/L2-capsids similarly bind to the surface of BPHE-1 cells. After incubation of the cells at 37C, the L1-capsids are diffusely distributed within the cytoplasm at 6 h; however, the L1/L2-capsids are aligned along distinct radial tracks across the cytoplasm and arrive in the perinuclear region within 2 h. Initially, L1/L2-capsids are located with cortical actin at the periphery of the cell and then are located with actin filaments.

Yang *et al.*[31] also showed that HPV16 L2 binds to b-actin. Bacterially expressed fusion protein comprised of GST/HPV16 L2



A B C

HPV16 13-ASATQLYKTCKQAGTCPPD-31 24-QAGTCPPDIIPKVEGKTIADQI–45 108-LVEETSFIDAGAP–120
HPV6 12-*******Q***LT******-30 23-LT******V*****HN******–44 107-*I**SAI*N****–119
HPV11 11-*******Q***AT******-29 22-AT******V*****HT******-43 106-*I**SAI*N****-118
HPV18 12-**V*D*******S******-30 23-QS******VV*****T*L**K*-44 106-*I*DS*VVTS***-118
HPV31 12-*******Q***A*****S*-30 24-A*****S*V***I*HT******-45 108-****SGIV*V***-120
HPV33 12-*******Q***AT******-30 23-AT******V******S******-44 107-*I******E****-119
HPV45 12-****D**R****S******-30 23-QS******V*N****T*L**K*-44 106-***DS*VVAS***-118
HPV52 12-*******Q***AS******-30 23-AS******V******T*****L-44 107-MI***T**ES***-119
HPV58 12-*******Q***AS******-30 23-AS******V******T******-44 107-*I**S********-119

Fig. 3. L2 regions binding to the cell surface. Region A: aa13 to 31.[34] Region B: aa24 to 45.[31] Region C: aa108 to 120.[33] Amino acid sequences of selected mucosal HPV L2s homologous to regions A, B, and C are aligned.

(aa1–128)/GFP is co-precipitated with b-actin in a lysate of radiolabeled SiHa cells, a human cell line derived from cervical cancer, with anti-GFP antibody. The co-precipitation is mediated by a complex formation between the HPV16 L2 region of aa25–45 with b-actin (Fig. 3, Region B). HPV16 pseudovirions containing L2 with a deletion of aa25–45 dramatically reduce infectivity, suggesting that the binding of L2 with b-actin is critical for infection.

We indicated that the HPV16 L2 region of aa108–126 plays an important role in the initial steps of infection.[32] The L2 peptide (aa108–126) tagged with GFP (L2/108–126/GFP) attaches to the HeLa cell surface at a low temperature and enters the cytoplasm when the temperature is raised to 37° (Fig. 3, Region C). L2/108–120/GFP does not bind to the cells pretreated with trypsin, suggesting that the cellular binding partner is a protein. Synthetic L2 peptide of aa108–120 interferes with the infectivity of the pseudovirions in COS-1 cells, suggesting the presence of a cell surface receptor for the aa108–120 region of L2, to which L2/108–126/GFP can bind. The aa substitutions that prevent L2/108–126/GFP from binding lower the infectivity of the pseudovirion having the same mutations. The data suggest that the N-terminal segment of L2 aa108–126,

displayed on the surface of capsids,[33] is involved in the early step of HPV infection by binding with the cellular receptor at the time of attachment of virions to cells.

L2 functions in the infectious process were also studied by combination of L2s of HPV16 and BPV1 by Yang *et al.*[34] The peptide with aa sequences of BPV1 aa1–88 binds to the surface of a variety of cells. The binding, which is not abolished by pretreatment of the cells with trypsin and heparinase, is interfered with by peptides of BPV1 L2 aa1–88 and HPV16 L2 aa13–31. HPV16 L2 aa13–31 peptide tagged with GFP (L2/13–31/GFP) attaches to HeLa cells (Fig. 3, Region A). Amino acid substitutions at aa18 plus aa19 and aa21 plus aa22 significantly reduce the ability of L2/13–31/GFP to bind to SiHa cells. Introduction of the same mutation into L2 reduces the infectivity of HPV16 pseudovirions. Deletion of the region of aa91–128 from BPV1 L2 eliminates infectivity of BPV1.

These studies indicate that interaction between L2 and the cell surface at a post-adsorption step would be important for HPV infection. The N-terminal segments of L2 aa13–31, aa25–45, and aa108–120 — which are displayed on the surface of capsids — may be involved in the early step of HPV infection by binding with the cellular molecules. The cellular target for the segment of aa25–45 is b-actin, but the cellular targets for aa13–31 and aa108–120 have not yet been identified.

Antigenicity of HPV Capsid Proteins

Conformationally Dependent and Linear Epitopes of L1-Capsids

The majority of anti-L1-capsid antibodies recognize conformationally dependent epitopes and bind to L1-capsids in a highly type-specific manner. Giroglou *et al.*[35] and Rose *et al.*[36] showed that anti-sera obtained by immunizing rabbits with HPV6, 11, 16, 18, 31, 33, 35, 39, and 45 L1-capsids are strongly reactive in ELISA with only the cognate L1-capsid type used as the immunogen. Similarly, we have showed that anti-sera obtained by immunizing mice with HPV6, 16, 18, and 58 L1-capsids react in ELISA with the cognate L1-capsid type used

as the immunogen.[37] The mouse antisera show low levels of cross-reactivity in ELISA using alkaline-denatured L1-capsids as the antigen.

HPV16 L1-capsids have conformationally dependent and linear epitopes. Christensen *et al.*[38] characterized a panel of anti-HPV16 L1 mouse monoclonal antibodies (MAbs) in detail. Fourteen MAbs were prepared by immunizing mice with HPV16 L1-capsids and examined for their reactivity with L1-capsids of HPV16, 6, 11, 18, 31, 33, 35, and 45. Three MAbs (H16.E70, H16.U4, and H16.V5) react with intact HPV16 L1-capsids in ELISA exclusively, indicating that type-specific conformationally dependent epitopes are displayed on the surface of the capsids. One MAb (H16.H5) reacts with both intact and denatured HPV16 L1-capsids and three MAbs (H16.C2, H16.L4, and H16.O7) react with denatured HPV16 L1-capsid, indicating that type-specific linear epitopes are either at the surface or buried within. One MAb (H16.P2) is weakly reactive to a shared conformationally dependent epitope present on L1-capsids of HPV16, 31, and 35. The other MAbs (H16.B20, H16.D9, H16.F18, H16.I23, H16.J4, and H16.S1) have various levels of cross-reactivity with the denatured L1-capsids of various HPVs.

Neutralization with Anti-L1 Antibodies in vitro

Giroglou *et al.*[35] showed that antisera obtained by immunizing rabbits with L1-capsids of HPV6, 11, 16, 18, 31, 33. 35, 39, and 45 neutralize the infectivity of pseudovirions in a type-specific manner, as predicted from the type-specific binding of antibodies against HPV L1-capsids.

Roden *et al.*[39] examined the neutralizing activity of the MAbs against HPV16 pseudovirions having BPV1 genome. After incubation with each MAb ascites sample to be tested, approximately 100 focus-forming units of pseudovirus were inoculated to mouse C127 cell monolayers, and the number of foci formed three weeks later was scored. Three type-specific and conformationally dependent MAbs (H16.E70, H16.U4, and H16.V5) inhibit focus formation with the pseudovirions, indicating that these MAbs have neutralizing activities *in vitro*. H16.D9, H16.H5, H16.J4, H16.L4, H16.O7 and H16.S1 do not neutralize the pseudovirions. H16.E70 and H16.V5

inhibit hemagglutination mediated by HPV L1-capsids or L1/L2-capsids, indicating that the MAbs prevented HPV pseudovirions from binding to the cell surface. H16.U4 recognizes an epitope in a functional domain not required for the binding (the role of the domain is not clear at present). Leucine residue at aa50 and serine residue at aa282 constitute part of the epitope for H16.V5 and H16.E70, respectively[39–41] (Fig. 1).

Wang *et al.*[42] showed that H16.V5 is capable of blocking the serological reactivity of the majority of human sera reactive to HPV16 L1/L2-capsids. ELISA plates coated with HPV16 L1/L2-capsids received an excess amount of each anti-HPV16 MAb, and then received human sera positive for anti-HPV16 antibodies. The pretreatment with H16.V5 prevents the majority of human sera from binding to the capsid antigen, suggesting that most human antibodies reactive with HPV capsids recognize the same or closely related major antigenic determinant(s) for H16.V5. The human sera that were not blocked by H16.V5 are mostly derived from the women with HPV16-associated cervical lesions.

Roden *et al.*[28] indicated that four anti-BPV1-L1 MAbs, which were produced by immunizing mice with purified BPV virions, inhibit the infection of mouse C127 cells with BPV1. These MAbs recognize type-specific conformationally dependent epitopes on the BPV1 L1-capsid. Three MAbs effectively block the binding of the virus to the cell surface at a ratio of approximately two MAb molecules per L1 molecule. One MAb does not block the binding effectively; however, it efficiently neutralizes BPV1 infectivity at 0.1 MAb molecule per L1 molecule.

Thus, anti-L1 antibodies can prevent papillomavirus infection *in vitro* by at least two steps: at the binding to the cell surface and at a currently unidentified subsequent step in the infection pathway.

Protective Immunity with Anti-L1 Antibodies in Animal Models

Consistent with the neutralization studies *in vitro*, Breiburd *et al.*[43] showed that antibodies recognizing conformationally dependent epitopes in L1 are required for protecting rabbits against challenge

with CRPV. Rabbits immunized with intact CRPV L1-capsids develop high-titer antibodies detectable by an ELISA using intact L1-capsids as antigen and are resistant to CRPV challenge. However, CRPV challenge to rabbits immunized with denatured CRPV L1/L2-capsids results in development of extensive papillomas. Transfer of serum or immunoglobulin G from rabbits immunized with CRPV L1-capsids protects naive rabbits against the challenge, strongly suggesting that the protection is mediated by neutralizing antibodies. Similarly, immunization of calves with L1- or L1/L2-capsids of BPV4 is extremely effective for protecting animals against BPV4 challenge.[44] The data obtained from these animal experiments clearly indicate that the properties of L1-capsids are suitable for prophylactic vaccine antigen.

Neutralization with Anti-L2 Antibodies in vitro

Roden *et al.*[45] showed that sheep antisera obtained by immunization with bacterially produced L2s of HPV6, 16, and 18 neutralize the pseudovirions of the homologous HPV type and cross-neutralize those of the heterologous HPV types, indicating that a cross-neutralization epitope(s) is present in L2.

We indicated the presence of a cross-neutralization surface epitope in HPV16 L2 by characterizing anti-HPV16 L2 MAbs.[21,33] Seventeen anti-HPV16 L2 MAbs were produced by immunizing Balb/c mice with HPV16 L1/L2-capsids. All of the MAbs bind HPV16 L1/L2-capsids in ELISA, indicating that they recognize epitopes displayed on the surface of L1/L2-capsids. ELISA using synthetic peptides as antigens showed that seven MAbs recognized linear epitope(s) within HPV16 L2 (473 amino acids) aa69–81, and two MAbs (#5 and #13) recognized those within aa108–120.[21,33] Amino acid sequences corresponding to the HPV16 aa108–120 are highly conserved among different HPVs (Fig. 3), and indeed MAb#5 and #13 bind to HPV6 L2.[33] MAb#5 and #13 inhibit infection with HPV16 and 6 pseudovirions of COS-1 cells, suggesting that the binding of antibody to the region of aa108–120 neutralizes the infectivity of HPV.[33]

We then showed that the peptide having the amino acid sequence of HPV16 L2 aa108–120 induces a cross-neutralizing antibody against

HPV16 and 6 in mice.[46] Balb/c mice were immunized nasally with the peptide, and the sera or vaginal washes were pooled from five animals for characterization. IgG antibodies in the sera and mucosal IgA antibodies in the vaginal washes bind to L1/L2-capsids of HPVs 6, 16, and 18, and neutralize HPV16 pseudovirions and HPV11 authentic virions.

Embers *et al.*[47] further characterized antibodies against HPV16 and HPV6 L2 distinct regions. The peptides having amino acid sequences of HPV16 L2 aa108–120 (16L2a), HPV16 L2 aa104–124 (16L2b: SIVSLVEETSFIDAGAPTSVP), HPV6/11 L2 aa108–120 (6/11L2a), and HPV6/11 L2 aa104–124 (6/11L2b:IVSLIEESAIINAGAPEIVPP) were synthesized and conjugated to keyhole limpet hemocyanin (KLH) for immunization or to bovine serum albumin (BSA) for ELISA. Antisera have been produced by immunizing rabbits and Balb/c mice (two animals per one immunogen) with one of the peptides. All of the sera react to the peptide used for immunization and show some level of cross-reactivity to the heterologous peptides. However, the antisera do not show significant cross-reactivity in ELISA using HPV16 and HPV11 L1/L2-capsids as antigens.

Neutralizing activities of the antisera were assessed by inhibition of infection with authentic HPV16 (obtained from cervical legions) and HPV11 virus (obtained from a condyloma and propagated by the use of human foreskin xenografts into nude mice). Neutralizing activity of the sera varied between two animals immunized with the same peptide in spite of the animals being inbred. Two rabbit sera, one obtained by immunization with 6/11L2a and one with 6/11L2b, have showed significant neutralizing activities to both HPV11 and HPV16 viruses. Rabbit antisera obtained by immunization with the HPV16 peptides showed very little ability to neutralize the HPV11 virus. Two mouse antisera obtained by immunization with H16L2a and one with H16L2b showed significant neutralizing activities to both HPV11 and HPV16 viruses. The H6/11L2a peptide was the poorest immunogen in mice for the production of neutralizing antibodies. The data support evidence that the epitope(s) in the L2 region containing aa108–120 may yield antibodies capable of neutralizing infectivity across viral types. The reason why the immunoreactions of the rabbits or mice vary with the individual rabbit or mouse is not clear.

HPV16 L2 N-terminal region, approximately aa18–144, is displayed on the surface of the L1/L2-capsid and the antibody capable of binding to the L2-surface region inhibits the infectivity of HPV pseudovirions. Kondo *et al.*[48] recently showed that rabbit antisera induced by the synthetic peptide with HPV16 L2 sequence aa18–38, 56–75, 61–75, 64–81, or 96–115 neutralize the HPV16 pseudovirions and cross-neutralize the pseudovirions of one or more of HPV18, 31 and 58. The data indicate that these antigen peptides contain cross-neutralization L2-epitopes.

Protective Immunity with Anti-L2 Antibodies in Animal Models

Chandrachud *et al.*[49] showed that immunization of cattle with bacterially expressed BPV4 L2 fused to glutathione S-transferase elicits its neutralizing antibodies and protects the animals against BPV4 challenge.

Embers *et al.*[50] showed that immunizations of rabbits with the peptides having amino acid sequences of rabbit oral (ROPV) and cutaneous (CRPV) papillomavirus L2 regions corresponding to HPV16 L2 aa108–120 protect the rabbits from challenge with ROPV and CRPV, respectively, further indicating that the epitope(s) required for protective immunity is located in the L2 region. The level of protection induced by these peptides is comparable to that seen with L1-capsid vaccination.

Prophylactic HPV Vaccine

Phase I/II clinical studies showed that an L1-capsid vaccine was well tolerable and highly immunogenic.[51–53] In the majority of the recipients, the serum antibody titers were higher than those detected in the natural infection.[51–53] A large-scale clinical trial in the United States is underway with an HPV16 prophylactic vaccine using HPV16 L1-capsids as an antigen.[1,53] Of women who received the vaccine, 99.7% were seroconverted. At 17.4 months after completion of the vaccination regimen, the incidence of persistent

HPV16 infection was 3.8 per 100 women-year at risk in the placebo group and 0 per 100 women-year at risk in the vaccine group, strongly suggesting that administration of the vaccine reduced the incidence of both HPV16 infection and HPV16-related cervical intraepitherial neoplasia. As predicted from type-specific neutralizing activity of anti-L1 antibodies, the cases of cervical intraepithelial neoplasia that were not associated with HPV16 infection were detected similarly among the placebo recipients and among the vaccine recipients. Thus, one of the remaining problems to be addressed is how to prevent infection with other high-risk HPVs.

We nasally administrated the synthetic peptide with the amino acid sequence of HPV16 L2 aa108–120 to ten healthy volunteers to evaluate the safety and immunogenicity of the peptide.[54] The administration of the peptide — 0.1mg to five recipients and 0.5 mg to five recipients — caused no serious local or systemic complications. The anti-L2 antibody binding to both HPV16 and L1/L2-capsids and neutralizing both HPV16 and 52 pseudovirions was induced in four of the five recipients in the 0.5 mg group. Serological responses were not induced by inoculation of 0.1 mg of the peptide. Although the number of the recipients is small, the data suggest the HPV16 L2 peptide is tolerable in humans and has the potential as a broad-spectrum prophylactic vaccine.

Conclusion

Immunization with HPV16 L1-capsids induces in rabbits and mice the antibodies recognizing conformation-dependent type-specific neutralization epitopes, whereas the antibodies induced by denatured L1-capsids do not neutralize HPV16 infectivity. A recent large-scale clinical trial showed that an HPV16 L1-capsid vaccine candidate was well tolerated by the participants and highly immunogenic in them. The results obtained so far indicated that the vaccine protected the recipients against HPV16 infection, which would progress to cervical cancer, in an HPV type-specific manner.

HPV16 minor capsid protein L2 has a linear neutralization epitope displayed on the surface of the L1/L2-capsids. The antibody

against this epitope seems to be capable of neutralizing the infectivity across the mucosal HPV types. The L2-epitope could be useful for the development of a prophylactic vaccine effective against a broader spectrum of HPVs associated with cervical cancer.

Acknowledgments

We thank Dr. Kunito Yoshiike for critical reading of the manuscript.

References

1. Koutsky LA, Ault KA, Wheeler CM, *et al.* (2002) A controlled trial of a human papillomavirus type 16 vaccine. *N Engl J Med* **347**: 1645–1651.
2. zur Hausen H. (1996) Papillomavirus infections — a major cause of human cancers. *Biochim Biophys Acta* 1288(2): F55–78.
3. zur Hausen H. (2002) Papillomaviruses and cancer: from basic studies to clinical application. *Nature Rev Cancer* **2**: 342–350.
4. Pfister H, Fuchs PG. (1994) Anatomy, taxonomy, and evolution of papillomaviruses. *Intervirology* **37**(3–4): 143–149.
5. Zhou J, Sun XY, Stenzel DJ, Frazer IH. (1991) Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is sufficient for assembly of HPV virion-like particles. *Virology* **185**: 251–257.
6. Hagensee ME, Yaegashi N, Galloway DA. (1993) Self-assembly of human papillomavirus type 1 capsids by expression of the L1 protein alone or by coexpression of the L1 and L2 capsid proteins. *J Virol* **67**: 315–322.
7. Kirnbauer R, Booy F, Cheng N, *et al.* (1992) Papillomavirus L1 major capsid protein self assembles into virus-like particles that are highly immunogenic. *Proc Natl Acad Sci USA* **89**: 12180–12184.
8. Kirnbauer R, Taub J, Greenstone H, *et al.* (1993) Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles. *J Virol* **67**: 6929–6936.
9. Rose RC, Bonnez W, Reichman RC, Garcea RL. (1993) Expression of human papillomavirus type 11 L1 protein in insect cells: *in vivo* and *in vitro* assembly of virus-like particles. *J Virol* **67**: 1936–1944.
10. Volpers C, Schirmacher P, Streeck R, Sapp M. (1994) Assembly of the major and the minor capsid protein of human papillomavirus type 33 into virus-like particles and tubular structures in insect cells. *Virology* **200**: 504–512.

11. Hagensee ME, Olson NH, Baker TS, Galloway DA. (1994) Three-dimentional structure of vaccinia virus-produced human papillomavirus type 1 capsids. *J Virol* **68**: 4503–4505.
12. Leder C, Kleinschumidt JA, Wiethe C, Mueller M. (2001) Enhancement of capsid gene expression: Preparing the human papillomavirus type 16 major structural gene L1 for DNA vaccination. *J Virol* **75**: 9201–9209.
13. Collier B, Oeberg D, Zhao X, Schwartz S. (2002) Specific inactivation of inhibitory sequences in the 5′ end of the human papillomavirus type 16 L1 open reading frame results in production of high levels of L1 protein in human epithelial cells. *J Virol* **76**: 2739–2752.
14. Oeben D, Collier B, Zhao X, Schwartz S. (2003) Mutational inactivation of two distinct negative RNA element in the human papillomavirus type 16 L2 coding region induces production of high level of L2 in human cells. *J Virol* 77: 11674–11684.
15. McCarthy MP, White WI, Palmer-hill F, *et al.* (1998) Quantitative disassembly and reassembly of human papillomavirus type 11 virus-like particles in vitro. *J Virol* **72**: 32–41.
16. Ishii Y, Yanaka K, Kanda T. (2003) Mutational analysis of human papillomavirus type 16 major capsid protein L1: the cysteins affecting the intermolecular bonding and structure of L1-capsids. *Virology* **308**: 128–136.
17. Chen XS, Garcea RL, Goldberg I, *et al.* (2000) Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Mol Cell* **5**: 557–567.
18. Trus BL, Roden RB, Greenstone HL, *et al.* (1997) Novel structural features of bovine papillomavirus capsid revealed by a three-dimensional reconstruction to 9 A resolution. *Nat Struct Biol* **4**: 413–420.
19. Sapp M, Fligge C, Petzak I, *et al.* (1998) Papillomavirus assembly requires trimerization of the major capsid protein by disulfides between two highly conserved cysteines. *J Virol* **72**: 6186–6189.
20. Zhou J, Sun XY, Louis K, Frazer IH. (1994) Interaction of human papillomavirus (HPV) type 16 capsid proteins with HPV DNA requires an intact L2 N-terminal sequence. *J Virol* **68**: 619–625.
21. Kawana K, Matsumoto K, Yoshikawa H, *et al.* (1998) A surface immunodeterminant of human papillomavirus type 16 minor capsid protein L2. *Virology* **245**: 353–359.
22. Roden RBS, Greenstone HL, Kirnbauer R, *et al.* (1996) *In vitro* generation and type-specific neutralization of a human papillomavirus type 16 virion pseudotype. *J Virol* **70**: 5875–5883.

23. Buck CD, Pastrana DV, Lowy DR, Schiller JT. (2004) Efficient intracellular assembly of papillomaviral vectors. **78**: 751–757.
24. Kawana K, Yoshikawa H, Taketani Y, *et al.* (1998) *In vitro* construction of pseudovirions of human papillomavirus type 16: Incorporation of plasmid DNA into reassembled L1/L2 capsids. *J Virol* **72**: 10298–10300.
25. Roden RBS, Kirnbauer R, Jenson AB, *et al.* (1994) Interaction of papillomaviruses with the cell surface. *J Virol* **68**: 7260–7266.
26. Mueller M, Gissmann L, Cristiano RJ, *et al.* (1995) Papillomavirus capsid binding and uptake by cells from different tissues and species. *J Virol* **69**: 948–954.
27. Volpers C, Unkell P, Schirmacher P, *et al.* (1995) Binding and internalization of human papillomavirus type 33 virus-like particles by eukaryotic cells. *J Virol* **69**: 3258–3264.
28. Roden RBS, Weissinger EM, Henderson DW, *et al.* (1994) Neutralization of bovine papillomavirus by antibodies to L1 and L2 capsid proteins. *J Virol* **68**: 7570–7574.
29. Giroglou T, Florin L, Schaefer F, *et al.* (2001) Human papillomavirus infection requires cell surface heparan sulfate. *J Virol* 75: 1565–1570.
30. Shafti-Karamat S, Handisurya A, Kriehuber E, *et al.* (2003) Different heparan sulfate proteoglycans serve as cellular receptors for human papillomaviruses. *J Virol* 77: 13125–13135.
31. Yang R, Yutzy IV WH, Viscidi RP, Roden RBS. (2003) Interaction of L2 with b-actin directs intracellular transport of papillomavirus and infection. *J Biol Chem* **278**: 12546–12553.
32. Kawana Y, Kawana K, Yoshikawa H, *et al.* (2001) Human papillomavirus type 16 minor capsid protein L2 N-terminal region containing a common neutralization epitopre binds to the cell surface and enters the cytoplasm. *J Virol* **75**: 2331–2336.
33. Kawana K, Yoshikawa H, Taketani Y, *et al.* (1999) Common neutralization epitope in minor capsid protein L2 of human papillomavirus types 16 and 6. *J Virol* **73**: 6188–6190.
34. Yang R, Day PM, Yutzy IV WH, *et al.* (2003) Cell surface-binding motifs of L2 that facilitate papillomavirus infection. *J Virol* 77: 3531–3541.
35. Giroglou T, Sapp M, Lane C, *et al.* (2001) Immunological analyses of human papillomavirus capsids. *Vaccine* **19**: 1783–1793.
36. Rose RC, Bonnez W, Rin CD, *et al.* (1994) Serological differentiation of human papillomavirus types 11, 16 and 18 using recombinant virus-like particles. *J Gen Virol* **75**: 2445–2449.

37. Matsumoto K, Yoshikawa H, Taketani Y, *et al.* (1997) Antibodies to human papillomavirus 16, 18, 58, and 6b major capsid proteins among Japanese females. *Jpn J Cancer Res* **88**: 369–375.
38. Christensen ND, Dillner J, Eklund C, *et al.* (1996) Surface conformational and linear epitopes on HPV-16 and HPV-18 virus-like particles as defined by monoclonal antibodies. *Virology* **223**: 174–184.
39. Roden RBS, Armstrong A, Haderer P, *et al.* (1997) Characterization of a human papillomavirus type 16 variant-dependent neutaralizing epitope. *J Virol* **71**: 6247–6252.
40. Slupetzky K, Shafti-Karamat S, Lenz P, *et al.* (2001) Chimeric papillomavirus-like particles expressing a foreign epitope on capsid surface loops. *J Gen Virol* **82**: 2799–2804.
41. White WI, Wilson SD, Palmer-Hill FJ, *et al.* (1999) Characterization of a major neutralizing epitope on human papillomavirus type 16 L1. *J Virol* **73**: 4882–4889.
42. Wang Z, Christensen ND, Schiller JT, Dillner J. (1997) A monoclonal antibody against intact human papillomavirus type 16 capsids blocks the serological reactivity of most human sera. *J Gen Virol* **78**: 2209–2215.
43. Breitburd F, Kirnbauer R, Hubbert NL, *et al.* (1995) Immunization with virus-like particles from cottontail rabbit papillomavirus (CRPV) can protect against experimental CRPV infection. *J Virol* **69**: 3959–3963.
44. Kirnbauer R, Chandrachud LM, O'Neil BW, *et al.* (1996) Virus-like particles of bovine papillomavirus type 4 in prophylactic and therapeutic immunization. *Virology* **219**: 37–44.
45. Roden RBS, Yutzy IV WH, Fallon R, *et al.* (2000) Minor capsid protein oh human genital papillomaviruses contains subdominant, cross-neutralizing epitopes. *Virology* **270**: 254–257.
46. Kawana K, Kawana Y, Yoshikawa H, *et al.* (2001) Nasal immunization of mice with peptide having a cross-neutralization epitope on minor capsid protein L2 of human papillomavirus type 16 elicit systemic and mucosal antibodies. *Vaccine* **19**: 1496–1502.
47. Embers ME, Budgeon LR, Culp TD, *et al.* (2004) Differential antibody responses to a distinct region of human papillomavirus minor capsid proteins. *Virology* **22**: 670–680.
48. Kondo K, Ishi Y, Ochi H, *et al.* (2007) Neutralization of HPV16, 18, 31, and 58 pseudovirions with antisera induced by immunizing rabbits with synthetic peptides representing segments of the HPV16 minor capsid protein L2 surface region. *Virology* **358**: 266–272.

49. Chandrachud LM, Grindlay GJ, McGarvie GM, *et al.* (1995) Vaccination of cattle with the N-terminus of L2 is necessary and sufficient for preventing infection by bovine papillomavirus-4. *Virology* **211**: 204–208.
50. Embers ME, Budgeon LR, Pickel M, Christensen ND. (2002) Protective immunity to rabbit oral and cutaneous papillomaviruses by immunization with short peptides of L2, the minor capsid protein. *J Virol* **76**: 9798–9805.
51. Evans TG, Bonnez W, Rose RC, *et al.* (2001) A Phase 1 study of a recombinant virus-like particle vaccine against human papillomavirus type 11 in healthy adult volunteers. *J Infect Dis* **183**: 1485–1493.
52. Harro CD, Pang YY, Roden RB, *et al.* (2001) Safety and immunogenicity trial in adult volunteers of a human papillomavirus type 16 L1 virus-like particle vaccine. *J Natl Cancer Inst* **93**: 284–292.
53. Schiller JT, Davies P. (2004) Delivering on the promise: HPV vaccines and cervical cancer. *Nature Rev Microbiology* **2**: 343–347.
54. Kawana K, Yasugi T, Kanda T, *et al.* (2003) Safety and immunogenicity of a peptide containing the cross-neutralization epitope of HPV16 L2 administered nasally in healthy volunteers. *Vaccine* **21**: 4256–4260.

Chapter 20

Chimeric Recombinant Hepatitis E Virus-like Particles Presenting Foreign Epitopes as a Novel Vector of Vaccine by Oral Administration

Yasuhiro Yasutomi[*,†]

Virus-like particles (VLPs) are useful for studies on virion formation, host immune responses to viruses, and vaccines in no practical cell culture systems to allow the growth of virus. Among the various non-replicating molecules, VLP, an empty particle with a structure similar to that of an authentic virus particle, offers the possibility of a new approach for these studies. Hepatitis E virus (HEV) is an unclassified calicivirus-like, positive-strand RNA virus that causes human acute hepatitis by fecal-oral transmission. HEV first infects epithelial cells of the small intestine and then reaches the liver through the portal vein. It has been reported that intact open reading frame 2 (ORF2) of HEV is expressed as a membrane glycoprotein when artificially expressed in mammalian cells *in vitro*, probably because the N-terminal amino acid sequence serves as a signal peptide.[1,2] On the other hand, the intact ORF2 expression in insect cells resulted in various sizes of proteins with cleavages on both N- and C-termini, of which 53 kD polypeptides secreted in the culture supernatant.[3–5] It has also been reported that only after cleavage of C-terminal, which results in molecular weight reduction to

*Laboratory of Immunoregulation and Vaccine Research, Tsukuba Primate Research Center, National Institute of Biomedical Innovation, Tsukuba, Ibaraki 305-0843, Japan.
†Department of Immunoregulation, Mie University Graduate School of Medicine, Mie 514-8507, Japan. E-mail: yasutomi@nibio.go.jp; yasutomi@doc.medic.mie-u.ac.jp.

54 kD, does ORF2 participate in the VLP formation.[6,7] In this chapter, investigation of HEV-VLP could be carried a molecule for foreign antigenic epitopes and to stimulate mucosal immunity without the need for adjuvant.

Chimerization of VLP

Chimerization of VLP is a unique and useful method for studying morphology, assembly and host recognition of a virus. However, there are not many reports about chimeric VLP in the fields of virology and immunology. Moreover, the induction of immune responses through natural infectious route against not only VLPs but also carrying epitopes is limited.[8–12] Some chimeric plant virus particles carrying foreign epitopes have been reported.[13] These chimeric virus particles are replication-competent and elicit immune responses through mucosal immunization. These particles were derived by foreign epitope insertion in a cDNA of a virus. Chimeric VLPs obtained by the same method have also been reported.[8–11] Successful chimerization is dependent on selection of an appropriate insertion site in VLPs. Another system for chimerization of VLP is co-infection of a couple of baculoviruses in the same cells, which allows VLPs to be obtained as chimeric VLPs.[12] This method is an easy way to obtain the chimeric VLP, although the stability of molecular constructs and the characteristics of morphology to original virus are not promised. Chimeric VLPs obtained by foreign-molecule insertion are suitable for studying morphogenesis of viruses and host recognition to both VLPs and inserted molecules.

HEV-VLP

Hepatitis E is an acute viral hepatitis caused by infection with HEV that was first recognized in India 1955. The HEV has been isolated from various animals, suggesting that hepatitis E is a zoonosis.[14,15] Although an *in vitro* culture system to amplify HEV has not been developed, over-expression of a part of ORF2 in a baculovirus expression system allows this protein to assemble into a VLP.[6] Formation of this VLP occurs only when N-terminal of ORF2 — where potential

signal sequence is encoded — was deleted from the expression construct.[6,16] It has also been reported that additional endogenous cleavage of 52 amino acids at the C-terminal is necessary for the assembly of a VLP.[6,7] HEV-VLP appears as an empty particle of a slightly smaller size than that of a mature HEV particle.[6,7] An HEV-VLP has several advantages for studying virus formation or host recognition. In our experience, large amounts can be easily obtained from standard cultivation protocols compared with amounts of other VLPs obtained. The amount of purified HEV-VLPs collected from culture supernatant of 50 to 100 μg/ml is more than 100 times greater than that of other VLPs. It has recently been found that the VLPs elicit strong immune responses when administrated orally into mice as same to a natural infection route.[17] Moreover, it has been shown that oral inoculation of cynomolgus monkeys with HEV-VLP prevents the infection of native HEV by intravenous injection.[18] These findings indicated that HEV-VLPs conserved original HEV construction to enter the target cells. Conservation of the virus construct in VLPs is very attractive for vaccines inducing the same type of immune responses to virus infection.

Chimeric HEV-VLP Carrying Foreign Epitope

pVL5480/7126, a baculovirus transfer vector that includes a portion of the ORF2 from HEV (dORF2), was described previously.[6] To insert the tag sequence within dORF2, oligonucleotides that encode the tag amino acid sequence were synthesized as shown in Table 1, and described previously.[9] The restriction sites used for insertion sites 1 to 4 were *Hind*III, *Sac*II, *Bss*HII, and *Sac*II sites at nucleotide positions 5679, 6245, 6664, and 6773, respectively. For each site, oligonucleotide pairs of Htg5(0) and Htg3(GA), Htg5(+1) and Htg3(0), Htg5(0) and Htg3(GG), and Htg5(+1) and Htg3(0) were used, respectively. A C-terminal tag was added at a position 52 amino acids upstream from the translational terminal. This site was chosen because the last 52 amino acids at the C-terminal of ORF2 are cleaved off during the formation of VLPs. The nucleotide sequences around the inserted tag are schematically shown in Fig. 1. The plasmid containing the chimeric dORF2 was co-transfected

Table 1. Oligonucleotides Used in this Study

Oligonucleotide	Sequence (5′ to 3′)
HEVBacBg	CGCAGATCTATGGCGGTCGCTCCAGCCC
HEV52Pr	CTGCAGCTATGCTAGCGCAGAGTG
Htg5(0)	CAGCCTGAACTCGCTCCAGAGGA
Htg5(+1)	GCCAGCCTGAACTCGCTCCAGAGGA
Htg3(0)	ATCTTCCGGATCCTCTGGAGCGAG
Htg3(GA)	TCATCTTCCGGATCCTCTGGAGCGAG
Htg3(GG)	CCATCTTCCGGATCCTCTGGAGCGAG
Tag(-52)	CTGCAGCTAATCTTCGGGGTCCTCCGGGGCGAGCT CAGGCTGTGCTAGCGCAGAGTGG
BglTag	AGATCTATGGCGCAGCCTGAACTCGCTCCAGAGGA TCCAGAAGATGCGGTCGCTCCAGCCCATGAC

N.(5486) ATGGCGCAGCCTGAACTCGCTCCAGAGGATCCAGAAGATGCGGTCGCT
M A Q P E L A P E D P E D A V A
Site 1.HindIII(5680) GAAGCTCAGCCTGAACTCGCTCCAGAGGATCCGGAAGATGAAGCTTCTAAT
E A Q P E L A P E D P E D E A S N
Site 2.SacII(6245) GGCCGCCAGCCTGAACTCGCTCCAGAGGATCCGGAAGATCGCGGGATA
G R Q P E L A P E D P E D R G I
(5480) N
dORF2
C (7126)
Site 3. Bss HII(6664) ACCGGCCAGCCTGAACTCGCTCCAGAGGATCCGGAAGATGGCGCGCAGGCC
G A Q P E L A P E D P E D G A Q A
Site 4.Sac II(6773) CTCCGCCAGCCTGAACTCGCTCCAGAGGATCCGGAAGATCGCGGTAAG
L R Q P E L A P E D P E D R G K
C.(6973) CTAGCACAGCCTGAGCTCGCCCCGGAGGACCCCGAAGATTAG
L A Q P E L A P E D P E D *

Fig. 1. Schematic diagram and sequences around the tag epitope insertion sites in dORF2. Upper rows show nucleotide sequences and lower rows show the corresponding amino acid sequences. Amino acid numbers relative to the full-length ORF2 are indicated next to the amino acid before the inserted amino acids. Nucleotide numbers referring to HEV genome are in parenthesis. Inserted sequences are underlined. The tag epitope amino acid sequence is in bold face.

with baculovirus DNA, and the recombinant baculovirus was generated as described previously.[6] The production and purification of HEV-VLP was performed as described previously.[6,7]

Utilizing any of the insertion sites, the chimeric dORF2 was expressed at almost equal levels in the cell lysates (Fig. 2A). The antigenicity of the tag epitope was maintained in all cases, as shown by

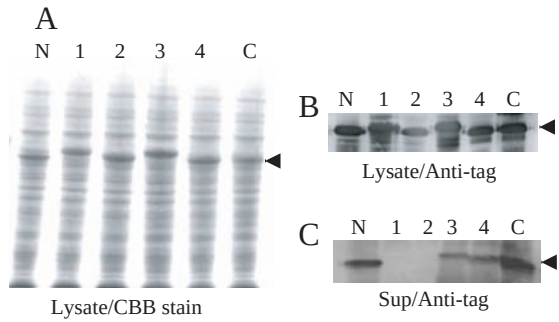


Fig. 2. Expression of chimeric dORF2. A. Expression in the cell lysates was examined by Coomasie brilliant blue staining. B. Antigenicity of the tag epitope in the cell lysates was confirmed by Western blotting with the anti-tag antibody. C. Presence of each chimeric dORF2 in cell supernatant (8 μl) was examined by Western blotting with the anti-tag antibody. The insertion site for each chimera is indicated at the top of the panel. N; N-terminal, 1 to 4; sites 1 to 4, respectively, C; C-terminal. The arrowhead on the right of each panel indicates the position of the chimeric dORF2.

Western blot analysis (Fig. 2B). Even at the C-terminal region, where the native HEV amino acid sequence is cleaved, the tag was not cleaved off from dORF2 (Fig. 2B, lane C). Among these chimeras, only the N- and C-terminal insertions resulted in release of a large amount of chimeric dORF2 into the culture supernatant (Figs. 2C and 2D), although small amounts were released when the insertions were made at either site 3 or 4. These results indicate that internal insertions somehow disturbed the release of dORF2 into the culture supernatant. The precise mechanisms involved in the HEV virion formation are not yet clear. The added tag at 52 amino acids upstream from the C-terminal region, where dORF2 is normally cleaved in insect cells, was not cleaved off in the infected cells during the generation of the chimeric VLPs. This is most likely due to alteration of the amino acid sequence recognized by the proteolytic enzyme involved in the C-terminal modification of HEV-VLP. The successful addition of extra amino acid sequences to the C-terminal of dORF2 suggests that the presence of extra amino acids at the C-terminal is not crucial

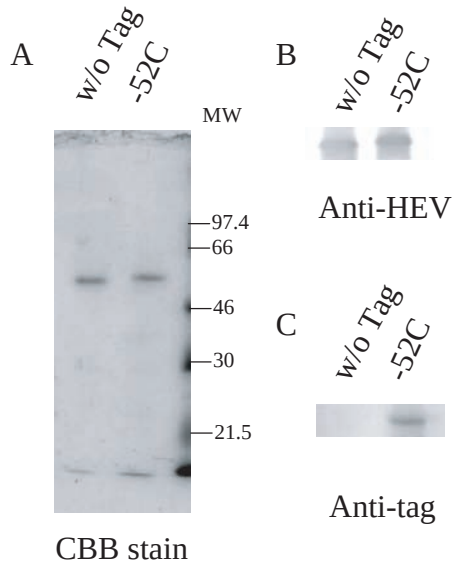


Fig. 3. Purification of the chimeric VLP. Purified VLP-52C was analyzed for its purity (A) and reactivity to anti-HEV (B) and anti-tag antibodies (C). A. Equal amounts (0.3 μg) of purified VLP-52C (–52C) and VLP without tag (w/o Tag) were separated on SDS-PAGE and stained by Coomasie brilliant blue staining. Positions of molecular weight markers are indicated on the right of the panel. B and C. Equal amounts (0.1 μg) of VLP-52C (–52C) and VLP without tag (w/o Tag) were analyzed by Western blotting using anti-HEV (B) and anti-tag (C) antibodies, respectively.

for preventing dORF2 from being incorporated into a VLP form. Rather, the amino acid sequences encoded by the HEV ORF2 genome prevented the formation of VLP.

We attempted to purify chimeric VLPs from the supernatant of Tn5 cells expressing chimeric dORF2 with a tag at either C-termini. The VLP-52C was slightly larger than the HEV-VLP without the tag (Fig. 3A). The purified VLP-52C retained the antigenicity of HEV as well as the intact tag epitope, as shown by the reactivity of specific antibodies (Figs. 3B and 3C).

Electron microscopic observation showed that VLP-52C was approximately 25 nm in diameter and indistinguishable from the

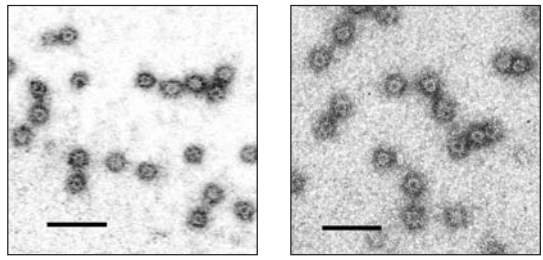


Fig. 4. Electron micrograph of VLP-52C. VLP-52C (A) and VLP without tag (B) were observed under electron microscopy after negative staining at a magnification of ×60,000. Inserted bar indicates 100 nm.

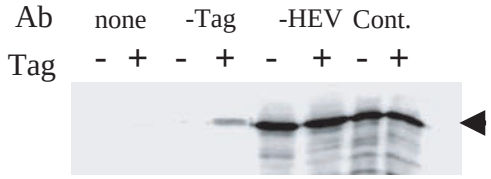


Fig. 5. Surface exposure of the tag epitope on VLP-52C. Surface exposure of the tag epitope on intact VLP-52C was examined by immunoprecipitation with the anti-tag antibody. Antibodies used are indicated at the top of panel. None, negative control without antibody; α-Tag, anti-tag antibody; α-HEV, anti-HEV antibody; cont., purified VLP-52C and VLP without tag were run as controls. The second row indicates either VLP with (+) or without the tag (–).

VLP without the tag (Fig. 4). Using two methods, we confirmed that the inserted epitope tag was exposed on the surface. The intact VLP-52C was immunoprecipitated with the anti-tag antibody, while the anti-HEV antibody immunoprecipitated both VLP-52C and the VLP without the tag (Fig. 5). Furthermore, the anti-tag antibody specifically reacted with the intact VLP-52C in an ELISA (data not shown). The results of immunoprecipitation and ELISA using intact chimeric VLP suggest that the tag epitope is exposed on the surface of the HEV-VLP.

Immune Responses to Chimeric HEV-VLPs through Oral Administration

Since many pathogenic viruses and bacteria establish their initial infections through the mucosal surface, vaccine strategies that can stimulate mucosal immunity have been widely studied (reviewed in Ogra *et al.*).[19] However, there are several difficulties in oral immunization with non-replicating molecules, such as low pH in the stomach, presence of proteolytic enzymes in the digestive tract, and presence of physical as well as biochemical barriers associated with the mucosal surface itself.[19] We previously reported that the HEV-VLP preserved original HEV construction and entered the epithelial cells of the small intestine by oral administration.[20] From these findings, mice were immunized with 50 μg of purified VLP-52C by the oral route four times at two-week intervals, with the mice having the ability to induce mucosal and systemic epitope-specific antibody responses. Specific IgG antibodies to the tag as well as to HEV were detected in intestinal fluids as early as two wpi (Fig. 6A). IgG levels in intestinal fluids continued to increase until the termination of the experiments. Specific IgA to both the tag and HEV also appeared in intestinal fluids from two wpi, paralleling the IgG levels (Fig. 6B). The IgA levels also continued to increase until the termination of experiments. As expected, the control mice immunized with VLP without the tag developed IgG and IgA only to HEV. In sera, levels of a specific IgG antibody to both the tag and HEV showed slightly higher OD values than those in non-immunized controls, but they never reached significant levels, as occurred in the intestinal fluids (Fig. 6C). The levels of specific IgA in sera were also low, although the OD values were also higher than the non-immunized controls (Fig. 6D). The control mice immunized with VLP without the tag showed similarly low OD values to HEV. The specific antibodies in the intestinal fluids were analyzed for their isotypes at 10 wpi. At this point, average OD values and SD for IgG and IgA in the intestinal fluids were 1.02±0.22 and 0.64±0.038, and 0.96±0.086 and 0.66±0.040, respectively, to HEV and the tag in three mice immunized with the chimeric VLP. In the control mice immunized with VLP without the tag, the average OD values and SD for IgG and IgA were 0.88±0.047 and 0.64±0.027, and 0.11±0.024 and 0.084±0.013, respectively, to HEV and the tag. All

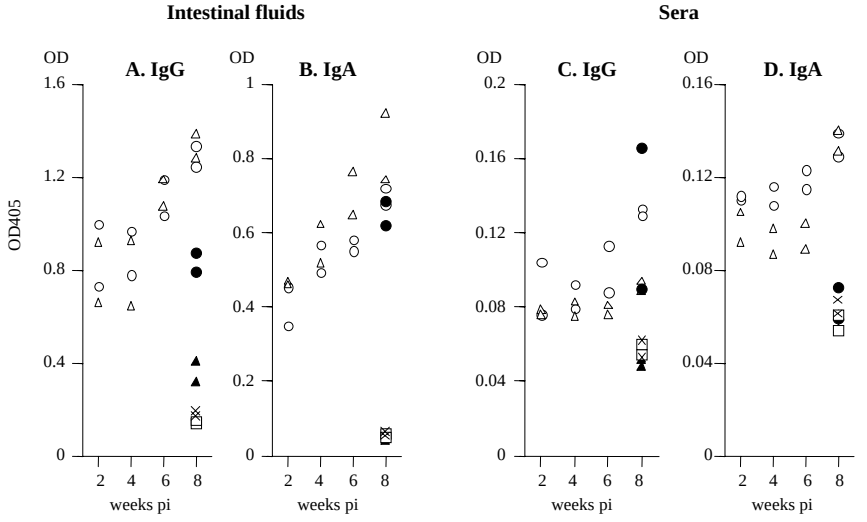


Fig. 6. IgG (A and C) and IgA (B and D) levels in intestinal fluids (A and B) and sera (C and D) of orally immunized mice. Circles and triangles indicate HEV-specific and the tag epitope-specific antibody levels, respectively, in individual mice. Two immunized mice were sacrificed at each time point (two, four, six and eight wpi). Specific antibody levels to HEV and the tag epitope of control mice immunized with VLP without the tag (closed circles and closed triangles, respectively) and background levels to HEV and the tag epitope of non-immunized mice (squares and crosses, respectively) are also shown. Antibody levels are indicated as OD405 in ELISA when sera and intestinal fluids were diluted at 1:100 and 1:2, respectively.

subclasses of IgGs to HEV — except IgG3, IgM, and IgA — were evident in all mice (Fig. 7B). Both to the tag and HEV, all mice failed to develop IgG3 above the detectable level (Figs. 7A and 7B). In the control mice immunized with VLPs without the tag, HEV-specific antibody reactions similar to the those with the chimeric VLPs were shown (Fig. 7B), while no detectable level of any isotype antibody specific to the tag was observed (Fig. 7A), as expected.

Induction of foreign epitope-specific antibody (Ab) responses by chimeric VLP administration is not easy compared with inducing cellular immune responses such as a cytotoxic T lymphocyte (CTL) response.[8,10,11,21,22] Moreover, our results showed Ab responses by oral

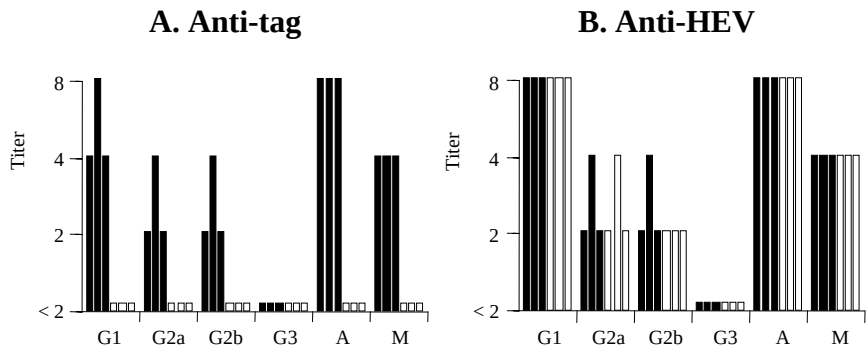


Fig. 7. Isotypes of antibodies specific to the tag epitope (A) and HEV (B) in intestinal fluids in orally immunized mice sacrificed at 10 wpi. Levels of IgA (A), IgM (M), and IgG subclasses (G1, G2a, G2b and G3) were examined by ELISA using isotype-specific secondary antibodies and are shown as end-point titers. Solid and open bars indicate antibody levels of each mouse immunized with the chimeric VLP and VLP without the tag epitope insertion, respectively.

administration to overcome the difficulties of a severe environment through the digestive tract. It is plausible that HEV-VLPs, which are derived from an orally transmissible virus, were incorporated into HEV-permissive epithelial cells in the small intestine because they retained structures and properties similar to those of HEV particles, producing an infection similar to that induced naturally.[17] It has been shown that the VLP structure should provide resistance to severe environments in the digestive tracts and enable specific binding to the mucosal surface if an appropriate VLP is chosen.[23] The delivery of a vaccine antigen (Ag) for induction of mucosal immune responses is usually achieved through the upper nasopharynx-associated lymphoid tissue (NALT), upper airway, salivary glands and tonsils.[24,25] Despite its obvious convenience, oral administration is rarely successful since it is quite difficult to protect vaccine Ag from the environment in the digestive tract.

The results of immunoprecipitation and ELISA using intact chimeric VLPs suggest that the tag epitope is exposed on the surface of the HEV-VLP. The successful induction of antibodies to the tag

also supports the hypothesis that the tag is exposed on the surface, since an internally localized B cell epitope in chimeric parvovirus-VLP failed to induce a specific antibody response.[11] Furthermore, this hypothesis is consistent with the results of three-dimensional analysis of the Norwalk virus-VLP particle, in which the C-terminal is exposed to the VLP surface.[26] Considering that B cell epitopes are generally hydrophilic and most likely exposed to the VLP surface, B cell epitope regions may not be directly involved in the protein-protein interactions to form a VLP. Our unsuccessful insertions into internal sites suggest that the integrity of internal regions must be maintained for proper protein folding and VLP formation. To find potential internal insertion sites, a precise three-dimensional structural map of the HEV-VLP may be necessary.

Oral vaccination has obvious advantages for a field trial in a large-scale public health vaccination program.[27] From a practical standpoint, oral administration is less stressful for vaccine recipients and does not require professional skill for administration. Moreover, delivery of vaccines via the intestinal tract is considered to be inherently safer than systemic injection. Encouraging results of phase I trials using Norwalk virus VLPs have recently been reported.[28] We also confirmed that chimeric HEV-VLPs carrying foreign CTL and B cell epitope at C-termini can elicit mucosal and systemic cellular immune responses as well as humoral immune responses by oral administration (submitted). It has become apparent that mucosal immune responses on different mucosal surfaces were achieved simultaneously, despite the initial stimulation of a single mucosal site.[29,30] Therefore, it is probable that oral administration of chimeric HEV-VLPs stimulates immune responses simultaneously on distant mucosal surfaces as well. This phenomenon significantly extends the potential use of chimeric HEV-VLPs as an oral vaccine vehicle.

A chimeric HEV-VLP has several advantages as an oral vaccine vector. First, large amounts can be easily obtained from standard cultivation protocols compared with amounts of other VLPs obtained. Second, the outcome of delivery of vaccine Ag in humans can be predicted using conventional laboratory animals, since HEV naturally infects various animals as well as humans through the same infectious

route and target cells.[6,31] Third, HEV-VLPs are stable at room temperature. Fourth, anti-HEV immune responses had no effect on boosting administration in the present study. Thus, HEV-VLPs are an attractive vaccine vector in developing countries because these VLPs can be preserved without the requirement of any particular equipment. These findings suggest that chimeric VLPs derived from orally transmissible viruses can be used as vaccine vectors to mucosal tissue by oral administration for the purpose of vaccination.

Acknowledgments

All experiments described in this chapter were carried out by Dr. Masahiro Niikura (Department of Microbiology and Molecular Genetics, Michigan State University) and Dr. Shiki Takamura (Department of Bioregulation, Mie University School of Medicine, Mie, Japan). This work was supported by Health Science Research Grants from the Ministry of Health, Labor and Welfare of Japan; the Ministry of Education, Culture, Sports, Science and Technology of Japan; and Regional Science Promotion Program.

References

1. Jameel S, Zafrullah M, Ozdener MH, Panda SK. (1996) Expression in animal cells and characterization of the hepatitis E virus structural proteins. *J Virol* **70**: 207–216.
2. Zafrullah M, Ozdener MH, Kumar R, *et al.* (1999) Mutational analysis of glycosylation, membrane translocation, and cell surface expression of the hepatitis E virus ORF2 protein. *J Virol* **73**: 4074–4082.
3. McAtee C, Zhang Y, Yarbrough PO, *et al.* (1996) Purification of a soluble hepatitis E open reading frame 2-derived protein with unique antigenic properties. *Protein Expr Purif* **8**: 262–270.
4. McAtee CP, Zhang Y, Yarbrough PO, *et al.* (1996) Purification and characterization of a recombinant hepatitis E protein vaccine candidate by liquid chromatography-mass spectrometry. *J Chromatogr B Biomed Appl* **685**: 91–104.
5. Robinson RA, Burgess WH, Emerson SU, *et al.* (1998) Structural characterization of recombinant hepatitis E virus ORF2 proteins in baculovirus-infected insect cells. *Protein Expr Purif* **12**: 75–84.

6. Li TC, Yamakawa Y, Suzuki K, *et al.* (1997) Expression and self-assembly of empty virus-like particles of hepatitis E virus. *J Virol* **71**: 7207–7213.
7. Xing L, Kato K, Li TC, *et al.* (1999) Recombinant hepatitis E capsid protein self-assembles into a dual-domain T = 1 particle presenting native virus epitopes. *Virology* **265**: 35–45.
8. Casal JI. (2001) Use of the baculovirus expression system for the generation of virus-like particles. *Biotechnol Genet Eng Rev* **18**: 73–87.
9. Niikura M, Takamura S, Kim G, *et al.* (2002) Chimeric recombinant hepatitis E virus-like particles as an oral vaccine vehicle presenting foreign epitopes. *Virology* **293**: 273–280.
10. Rueda P, Hurtado A, del Barrio M, *et al.* (1999) Minor displacement in the insertion site provokes major differences in the induction of antibody responses by chimeric parvovirus-like particles. *Virology* **263**: 89–99.
11. Rueda P, Marinez-Torrecuadrada JL, Sarraseca JC, *et al.* (2000) Engineering parvovirus-like particles for the induction of B-cell, CD4+ and CTL responses. *Vaccine* **18**: 325–332.
12. Yao Q, Zhang R, Guo L, *et al.* (2004) The cell-independent immune responses to chimeric hemagulutinin/simian human immunodeficiency virus-like particles vaccine. *J Immunol* **173**: 1951–1958.
13. Modelska A, Dietzschold B, Sleysh N, *et al.* (1998) Immunization against rabies with plant-derived antigen. *Proc Natl Acad Sci USA* **95**: 2481–2485.
14. Meng XJ, Purcell RH, Halbur PG, *et al.* (1997) A novel virus in swine is closely related to the human hepatitis E virus. *Proc Natl Acad Sci USA* **94**: 9860–9865.
15. Nishizawa T, Takahashi M, Mizuo H, *et al.* (2003) Characterization of Japanese swine and human hepatitis E virus isolates of genotype IV with 99% identity over the entire genome. *J Gen Virol* **84**: 1245–1251.
16. Tam AW, Smith MM, Guerra ME, *et al.* (1991) Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. *Virology* **185**: 120–131.
17. Li TC, Takeda N, Miyamura T. (2001) Oral administration of hepatitis E virus-like particles induces a systemic and mucosal immune response in mice. *Vaccine* **19**: 3476–3484.
18. Li TC, Suzaki Y, Ami Y, *et al.* (2004) Protection of cynomolgus monkeys against HEV infection by oral administration of recombinant hepatitis E virus-like particles. *Vaccine* **22**: 370–377.

19. Ogra PL, Faden H, Welliver RC. (2001) Vaccination strategies for mucosal immune responses. *Clin Microbiol Rev* **14**: 430–435.
20. Takamura S, Niikura M, Li TC, *et al.* (2004) DNA vaccine-encapsulated virus-like particles derived from an orally transmissible virus stimulates mucosal and systemic immune responses by oral administration. *Gene Ther* **11**: 628–635.
21. Morton G, Rueda P, Sarraseca JC, *et al.* CD8alpha-CD11b$^+$ dendritic cells present exogenous virus-like particles to CD8$^+$ T cells and subsequently express CD8 alpha and CD205 molecules.
22. Rueda P, Morton G, Sarraseca JC, *et al.* (2004) Influence of flanking sequence on presentation efficiency of a CD8$^+$ cytotoxic T-cell epitope delivered by parvovirus-like particles. *J Gen Virol* **85**: 563–572.
23. Ulrich R, Nassal M, Meisel H, Kruger DH. (1998) Core particles of hepatitis B virus as carrier for foreign epitopes. *Adv Virus Res* **50**: 141–182.
24. McCluskie MJ, Chu Y, Xia JL, *et al.* (1998) Direct gene transfer to the respiratory tract of mice with pure plasmid and lipid-formulated DNA. *Antisense Nucleic Acid Drug Dev* **8**: 401–414.
25. Morrow CD, Novak MJ, Ansardi DC, *et al.* (1999) Recombinant viruses as vectors for mucosal immunity. In J.-P. Kraehenbuhl & M.R. Neutra (eds), *Current Topics in Microbiology and Immunology* 236, pp. 255–273. Springer-Verlag, Berlin.
26. Prasad BV, Hardy ME, Dokland T, *et al.* (1999) X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**: 287–290.
27. Nagaraj K, Babu BV. (1997) Field trials of oral cholera vaccine in Vietnam. *Lancet* **349**: 1253–1254.
28. Ball JM, Graham DY, Opekun AR, *et al.* (1999) Recombinant Norwalk virus-like particles given orally to volunteers: phase I study. *Gastroenterology* **117**: 40–48.
29. Boyaka PN, Marinaro M, Vancott JL, *et al.* (1999) Strategies for mucosal vaccine development. *Am J Trop Med Hyg* **60**: Suppl., 35–45.
30. Mestecky J, McGhee JR. (1987) Immunoglobulin A (IgA): molecular and cellular interactions involved in IgA biosynthesis and immune response. *Adv Immunol* **40**: 153–245.
31. Nahde T, Muller K, Fahr A, *et al.* (2001) Combined transductional and transcriptional targeting of melanoma cells by artificial virus-like particles. *J Gene Med* **3**: 353–361.

Chapter 21

Nucleocapsid Protein of Hantaviruses (*Bunyaviridae*): Structure and Functions

*Alexander Plyusnin**, *Vibhor Kumar*†, *Olli Vapalahti*‡, *Peter Engelhardt*§

Hantaviral N protein is a typical nucleocapsid protein of a negative-strand virus. It is a multifunctional molecule that plays an important role in the viral RNA encapsidation and replication, and also in the virus assembly. In this chapter we review the current knowledge on hantaviral N protein structure and functions. We also present new data on a 3D structure of the hantaviral N-ptotein trimer and discuss its possible role in the viral replication cycle. Finally, we describe the structure of the N protein oligomers and polymers.

Introduction

Hantaviruses (genus *Hantavirus*, family *Bunyaviridae*) are negative-strand RNA viruses (NSRV) with a tripartate genome, each carried by

*Department of Virology, Haartman Institute, University of Helsinki, P.O. Box 21 (Haartmaninkatu 3), FI-00014 Helsinki, Finland; E-mail: alexander.plyusnin@helsinki.fi

†Laboratory of Computational Engineering, Helsinki University of Technology, P.O. Box 9203 (Innopolis 2, Tekniikantie 14), FI-02015 Espoo, Finland; E-mail: vkumar@cc.hut.fi

‡Department of Virology, Haartman Institute, University of Helsinki, P.O. Box 21 (Haartmaninkatu 3), and Department of Basic Veterinary Sciences, Faculty of Veterinary Medicine P.O. Box 66 FIN-00014, University of Helsinki and HUSLAB, Helsinki, Finland; olli.vapalahti@helsinki.fi

§Laboratory of Computational Engineering, Helsinki University of Technology, P.O. Box 9203 (Innopolis 2, Tekniikantie 14), FI-02015 Espoo, Finland; and Department of Pathology and Virology, Haartman Institute, University of Helsinki, P.O. Box 21 (Haartmaninkatu 3), FI-00014 Helsinki; peter.engelhardt@helsinki.fi

a specific, chronically infected rodent host. Some hantaviruses are human pathogens causing acute, and sometimes fatal, diseases known as Haemorrhagic Fever with Renal Syndrome and Hantavirus (Cardio) Pulmonary Syndrome, while others are apathogenic.[1] Hantaviral genome consists of three segments — Small (S), Medium (M) and Large (L) — encoding respectively the nucleocapsid (N) protein, two surface glycoproteins (Gn and Gc), and the viral polymerase (L protein). Hantaviral particles are spherical or oval in shape and approximately 80–120 nm in diameter. They have an envelope that is formed by membrane lipids derived from a host cell and spiked with Gn-Gc heterodimers. Three genome RNA (vRNA) segments are encapsidated by the N protein and form separate, filamentous ribonucleoproteins (RNPs). The L protein is associated with these RNPs, presumably via binding to panhandle-like structures formed by 3′- and 5′-terminal stretches of nucleotides that are complementary to each other. Bunyaviruses do not possess a membrane protein and it is thought that the viral RNPs are attached to the inner membrane via interaction(s) with the cytoplasmic tail of Gn protein.[1]

Although some details of the hantavirus replication cycle remain unclear, the general scheme seems to be the following (Fig. 1) (for a review, see Kukkonen *et al.*, 2005[2]). Replication takes place in a cytoplasm of infected cells that hantaviruses enter, after attachment to beta3 or beta1 integrins,[3] using the clathrin-mediated pathway,[4] although other receptors are also likely to be involved. In early endosomes, an acidic environment induces a fusion of viral and endosomal membranes and the viral RNPs are released into cytoplasm. For primary transcription, hantaviruses use the cap-snatching mechanism: the L protein cleaves 7–18 nucleotide-long primers from the 5′ termini of cellular mRNA molecules. Transcription then continues according to the "prime-and-realign" mechanism[5] without polyadenylation of mRNAs. Without delay, the virus-specific mRNAs are translated into viral proteins and, after reaching a threshold level of these proteins in the cytoplasm, the replication of viral genome starts. First, complementary RNA (cRNA) is synthesized; this RNA is used to produce new vRNA, which, in turn, activates secondary transcription.[6] The newly synthesized cRNA

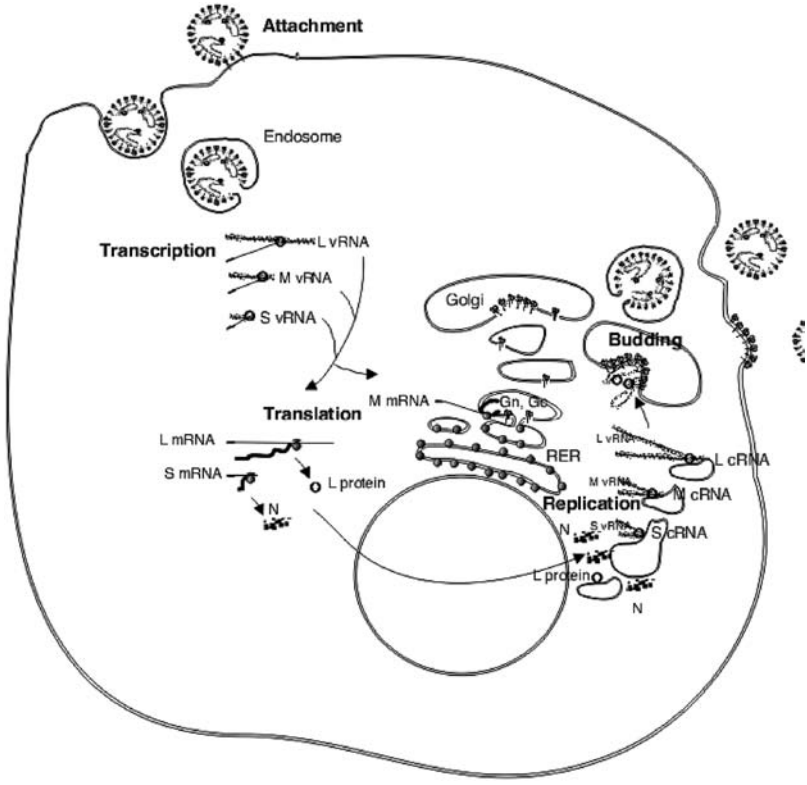


Fig. 1. Hantaviral replication cycle.

and vRNA are immediately encapsidated by the N protein. Gn and Gc proteins are translated from the same M mRNA, translocated to ER, cotranslationally cleaved by signal peptidases, glycosylated, and transported to the Golgi complex or to the plasma membrane, the places of hantavirus maturation and budding. Three viral RNPs, together with the L protein, become included in the budding enveloped particles as well, and the first hantavirus particles are released within 24 h p.i.[7]

Like nucleoproteins of other NSRVs, hantavirus N protein is a multifunctional molecule with both housekeeping and ambassadorial duties (for a review, see Kaukinen *et al.*, 2005[8]). It plays a crucial role in the viral RNA replication and encapsidation and also in the virus

assembly. In addition, it can modulate metabolism in infected cells via interactions with several cellular proteins, e.g. with Daxx- and SUMO-1-pathway members. In this chapter we review the current knowledge on hantaviral N protein structure and functions. We also present new data on a 3D-structure of the hantaviral N-ptotein trimer and discuss its possible role in the viral replication cycle. Finally, we describe the structure of the N protein oligomers and polymers.

General Features of the N Protein

The N protein of hantaviruses contains 429 to 433 aa residues. The protein sequence is well conserved within a given hantavirus type and, to a lesser extent, among all hantaviruses. Analysis of the sequence conservation reveals three conserved regions separated by two more variable regions spanning aa 50–80 and 230–310.[8] The central conserved region contains a cluster of positively charged residues that overlaps with a putative RNA-binding domain.[9]

Formation of RNPs, which are functional templates for transcription and replication of the viral genome, depends on interactions of the N protein with several partners: vRNA, other N protein molecules, and, perhaps, the L protein. Terminal panhandle-forming regions of the vRNA molecules seem to possess a unique binding region for the N protein.[10–12] The corresponding sequences of the cRNAs are recognized with lower affinity. Recently solved 3D structures of the N proteins of lyssa- and rhabdoviruses[13,14] implied a conformational change of the nucleocapsid protein molecule, which is essential for RNA encapsidation. Most probably, this conformational change results from interaction not only between the N protein and viral RNA but also between two (or more) molecules of the N protein: to-be-recruited RNA-binder and neighboring molecule(s) that have been already engaged with RNA and have adopted a new conformation. This model is in agreement with experimental data on *in vitro* binding of recombinant GST-N-trimer to the viral RNA panhandle.[12] Binding of the N-trimer to the panhandle might have important consequences. According to the model suggested by Mir and Panganiban,[15] the trimer can work as an RNA chaperone causing dissociation of

the panhandle and thus making the 3′ terminus of RNA available for the viral polymerase.

The phenomenon of hantaviral N protein oligomerization was discovered simultaneously by research groups from Finland[16] and the United States.[17] It was found that the N protein can form stable trimers in both viral particles and infected cells, and it was suggested that these trimers represent intermediates in the process of N protein oligomerization and RNP formation. An interaction between N protein monomers appeared to be non-covalent and electrostatic.[16] The studies that followed have further characterized the interacting domains and pinpointed aa residues that are crucial for the interaction.[18–21] The oligomerization of the hantaviral N protein was analyzed using a combination of computer modeling with the following *in vitro* techniques: peptide and alanine scanning, a mammalian two-hybrid assay, and an immunofluorescence assay. It was shown that both termini of the molecule contribute to the N-N-interaction. The data supported a "head-to-head, tail-to-tail" model, which suggests that the N-N-interaction is a two-step process involving an initial interplay between the N-terminal domains followed by a consolidating interaction between the C-terminal domains. The N-terminal tri/oligomerization domain of the hantaviral N protein was found to fold into a coiled-coil structure.[17,21,22] The results obtained were consistent with the existence of an anti-parallel coiled-coil, and several aa residues (L44, L58 and I51), all located at the "a" positions of heptad repeats, were found to be especially important for the N-N-interaction.[21] The consolidating interaction between the C-terminal domains were found to occur via protrusions of two alpha-helices and the shared hydrophobic space formed by amino acid residues 380-IILLF-384 and 413-LI-414.[19] All hantaviruses seemed to have the same, or very similar, structure of the terminal domains, which suggested a genus-specific mode of the N protein tri/oligomerization.

Experiments with *Hantaviral minigenomes* demonstrated that functional L and N proteins are the only viral proteins needed for replication and transcription.[23] Both proteins are at least partially colocalized at the Golgi compartment[24] and most probably interact with each other in determining the location of RNA synthesis of hantaviruses

and in directing RNPs to the site of virus assembly and budding. The tail of the Gn protein located on the cytoplasmic side of the Golgi membrane could also determine the RNPs that are included in the virion. It has been suggested that the cytoplasmic tail of Gn interacts with the N protein.[25] This could be an important step in virus assembly, which would secure the incorporation of RNP into virus particles.

The hantaviral N protein seems to interact with a variety of cellular proteins (for a review, see Kaukinen *et al.*, 2005[8]), but the discussion of these activities would go beyond the scope of this chapter.

3D-Structure of the N Protein

Crystallization of hantavirus N protein is pending. Purification of even modest quantities of the protein suitable for crystallization (i.e. water-soluble) is difficult to accomplish from purified virus particles because hantaviruses grow to low titers in cell cultures. As for recombinant N protein produced in heterologous expression systems (e.g. in the baculovirus-driven system), it appears to be "sticky" and tends to form aggregates that can be solubilized in 6–7 M urea, for example, but not directly in water or aquatic buffers.[26] Luckily, a substantial portion of the hantavirus N protein expressed in a baculovirus-based system exists in the form of a stable trimer, in which monomers are covalently linked to each other via S-S-bridges between cystein residues.[16] This trimer presented a useful model for our initial structural studies.

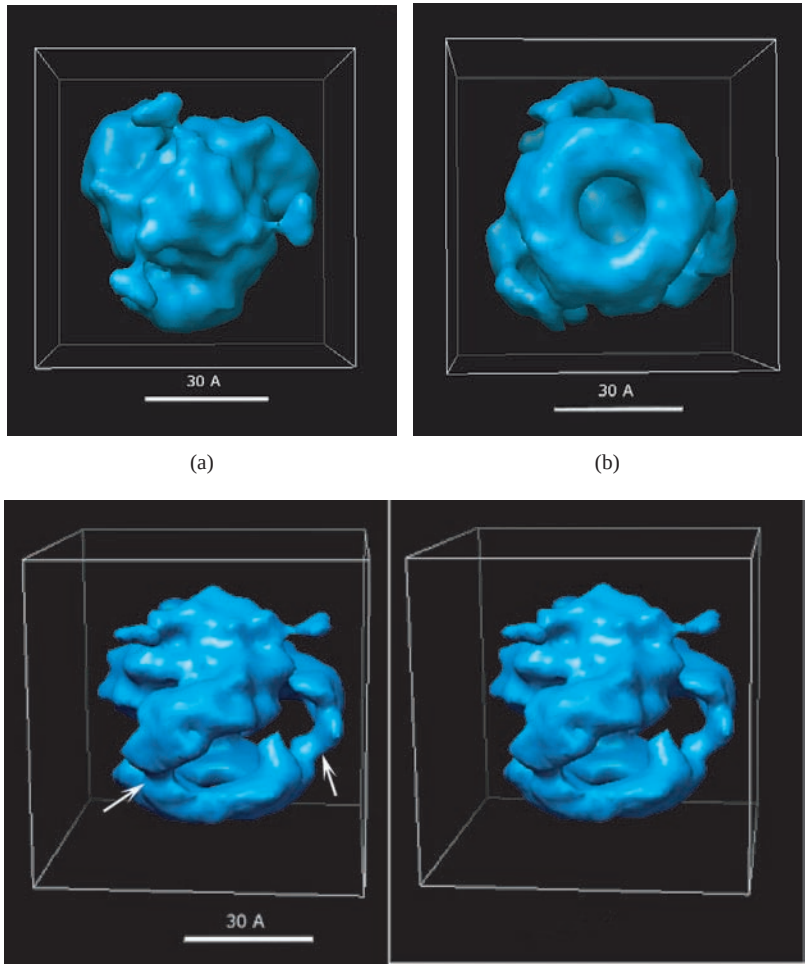
First, cryo-EM analysis of recombinant N protein was performed.[27] It confirmed the existence of the N-trimer and suggested a curved shape for the N protein monomers (with dimensions of approximately 8 nm by 3 nm), which resemble influenza and rabies virus N protein monomers in the reconstructed 3D models.[28,29] Importantly, contacts on both N- and C-termini of the monomers were clearly seen. In agreement with these data, the 3D reconstruction of the N-trimer using EM after negative staining revealed curve-shaped monomers attached to each other at both ends.[19] These observations, together with results of *in silico* docking of monomers into a trimer, were helpful in the mapping of oligomerization domains.[19,21]

To reconstruct the N-protein trimer in greater detail, the single-particle reconstruction was done assuming the symmetry of three.

With this assumption, the reconstructed shape of the N-trimer (Fig. 2) appeared slightly different from the previously obtained structure,[19] probably due to "averaging out" non-symmetrical parts of the monomers. Most importantly, certain structural details that were not seen in the earlier reconstruction became more dominant and clear. One such detail is the whirl-like twisted structure formed by the N-terminal domains of the monomers (Fig. 2a). The tip of this structure is probably formed by the coiled-coil helices.[21] Another detail is the ring-like structure formed by the C-terminal alpha-helices of the monomers (Fig. 2b). A close similarity was apparent between this ring and a structure formed by the nucleocapsid protein of rabies virus (cf. Fig. 4 in Schoehn *et al.*, 2001).[29] Via its hydrophobic amino acids, the ring formation of the hantaviral N-trimer might interact with a cellular membrane maintaining proper localization of the protein and, perhaps, might also facilitate interaction(s) with the cytoplasmic tail of the membrane-associated Gn protein.

It should be noted that what has been accomplished using a single three-fold axis of rotational symmetry was an idealized reconstruction of the N-trimer. We cannot rule out a possibility that the trimer is not absolutely symmetrical and that the monomers could be arranged in a slightly different way (e.g. in the slider of a zipper, one gate differs from the other two gates). Consequently, to assure a proper functioning of the N-trimer (see below), the cavities in the N-monomers (= gates) can acquire somewhat different shapes. In-depth structural studies of the N-trimers (using non-symmetry) are needed to clarify this issue; an increased resolution will be crucial to perform this task.

From basic docking tests (with SITUS) we have developed some ideas on how the N-trimer could interact with the viral RNA. These interactions would be crucial for the suggested RNA chaperoning activity of the N protein, i.e. for the formation and dissociation of the RNA-panhandle.[15] Figures 3 and 4 show the N-trimer with *in silico* docked double-stranded RNA (ds RNA) that imitates the terminal panhandle structure of vRNA (cRNA) and also single-stranded RNA (ssRNA) that imitates vRNA/cRNA outside the panhandle (or of the melted panhandle). When an inverted electron density map of N-protein trimer has been used (to represent the hollow space in and around the trimer), ssRNA or dsRNA were docked and regularly



(a) (b)

(c)

Fig. 2. 3D reconstruction of the hantaviral N-protein trimer performed using a single particle reconstruction. The reconstruction was done assuming symmetry (C3), i.e. single three-fold axis of rotational symmetry. 3 μL sample of purified recombinant Puumala virus N protein[26] (0.5–1 mg/mL, dissolved in 6 M urea, 10 mM Tris-HCl, pH 8.0) was applied to carbon film-coated 300- or 400-mesh-Au grids (Quantifoil), diluted and washed, i.e. floated in sequence in drops of 1% uranyl acetate and negatively stained. Electron microscopic pictures were collected at magnification of 50,000 at 80 kV with a Jeol 1200EX microscope. Negatives were scanned at 4,000

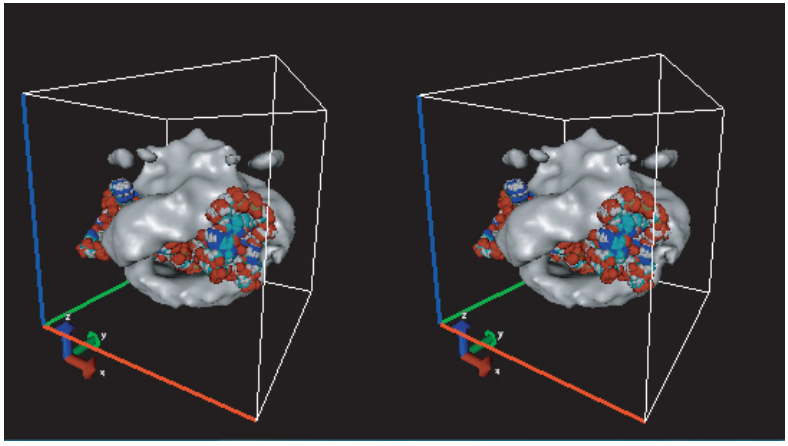
fitted in the same three cavities in the N-trimer (Fig. 3). However, when a non-inverted electron density map of N-trimer was used, both RNA types preferred an area near the N-terminal domains of three monomers (Fig. 4). Such a location might be of special importance as it has been suggested that the N-terminus carries an RNA chaperone domain.[15] The docking of the 12 nt long ssRNA or 20 nt long dsRNA into the trimer plainly shows that the RNA spans more than one N-monomer. Thus the RNA chaperone activity of the N protein would require the trimer structure since single monomers apparently cannot wrap around the RNA-panhandle or the panhandle-forming terminal RNA sequences due to size restrictions.

We hypothesize that the N-trimer mediates panhandle dissociation in the following way:

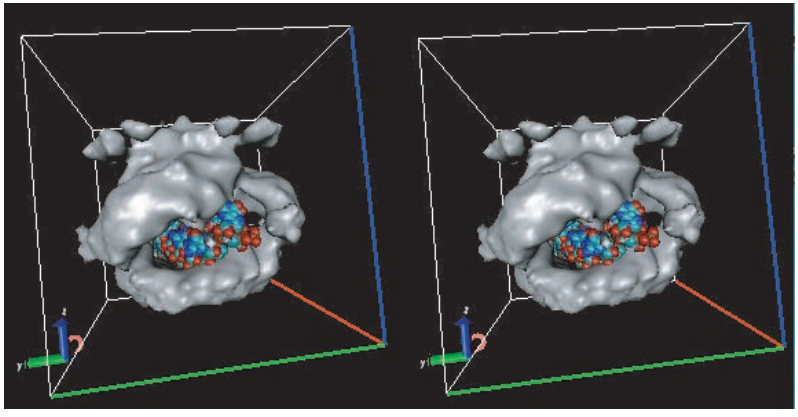
1. The process is started by the trimerized coil-coiled helices of the N-terminal domain of the N protein, which recognize the dsRNA panhandle and attach to it.
2. The N-trimer starts to move along the panhandle, thus the end of the panhandle goes into a gap between the middle domains of

---

dots per inch (Nikon LS-8000 ED). The random extensive sampling method for indistinctive protein particles from the electron micrographs (for the reference, see Kaukinen *et al.*, 2004[19]) was used for picking projections (~38,000) of the trimers. The top view (a) exposes the twisted structure presumably formed by the N-terminal coiled-coil helices of the monomers.[21] The bottom view (b) exposes a ring formation in the C-terminal region of the trimer presumably formed by interacting alpha-helices,[19] in analogy with the nucleocapsid protein of rabies virus (cf. Fig. 4 in Schoehn *et al.*, 2001[29]). The two images in (c) show a stereo pair of tilted side views, showing part of three symmetrically located connected cavities. The 3D reconstruction was done with EMAN (http://ncmi.bcm.tmc.edu/%7Estevel/EMAN/doc/index.html) and visualized with CHIMERA (http://www.cgl.ucsf.edu/chimera/). Note the twisted appearance of the N-monomers and the relatively thin junction between their N- and C-terminal domains (arrows), which parallel the X-ray crystallography data on the N proteins of rabies and vesicular stomatitis viruses (cf. Figs. 1 and 3 in Albertini *et al.*, 2006[13] and Green *et al.*, 2006,[14] respectively).

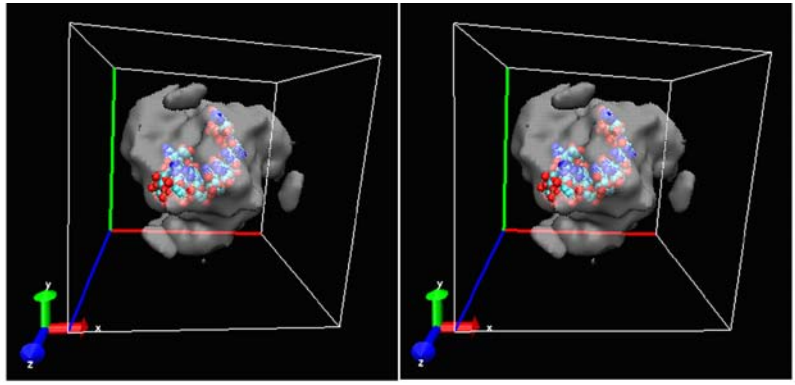


(a)

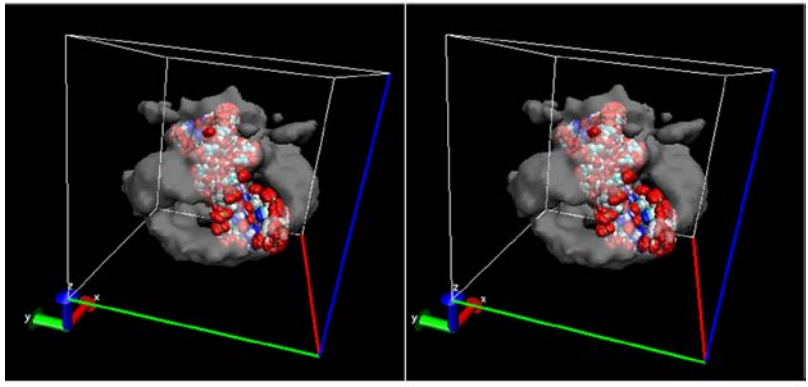


(b)

Fig. 3. Stereo pairs of tilted side views of the reconstructed N-trimer showing the docking of three double-stranded RNA (dsRNA), each 20 nucleotides (nt) long (**a**), and three-single stranded RNA (ssRNA), each 12 nt long (**b**) to the three symmetrically joint cavities of the trimer. The preferred locations for dsRNA and ssRNA in the trimer were found by performing the RNA docking with an inverted 3D electron density map of the trimer (inverted 3D reconstructions would represent the hollow space in and around the trimer). SITUS was used for docking (http://situs.biomachina.org/), and VMD was used for visualization of stereo pairs of the docked RNAs with the density maps of N-protein as suggested in SITUS.



(a)



(b)

Fig. 4. Stereo pairs of tilted top (**a**) and side (**b**) views of docking ssRNA and dsRNA to electron density maps of non-inverted N-trimer (using SITUS and visualized with VMD, see Fig. 3). Note that the N-trimer is shown as transparent.

the monomers and further inside the internal cavity of the trimer, as shown in Figs. 3 and 4.

3. Once entering the trimer, the dsRNA end cannot freely come out from one of the two remaining gaps because of the steric hindrance, which is presented by the monomer located just in front of the entrance hole marked by arrows on Fig. 2. But,

since the N-terminal of the trimer continues to move ahead and push dsRNA even further inside the cavity, the steric conflict is resolved by melting the dsRNA.

4. The 3′ terminus of the vRNA comes out of the trimer where it is immediately engaged by the viral polymerase, while the 5′ terminus remains captured by the N-trimer.

The scheme presented here has an analogy in the mode of activity of cellular DNA topoisomerase.[30] Coiled-coil helices of this enzyme first recognize dsDNA; the protein then moves along causing dsDNA to move down to other parts of the protein where they are opened or chopped. Our observation on high sequence (and hence structural) homology between the N-terminal, coiled-coil-forming region of the hantaviral N protein and DNA topoisomerase[21] is in line with this reasoning. There is also an example of an RNA chaperone (the L1 protein of the mammalian retrotransposon) that acts as a stable trimer formed by means of the N-terminal coiled-coils.[31]

Figure 5 shows how the N protein could encapsidate RNA (in this case, RNA from insect cells infected with a recombinant baculovirus expressing the N protein of Puumala hantavirus). The image on the background shows RNP complexes, which appear as dark lines surrounded by bright thick boundaries on both sides. A framed part of the RNP is shown in detail at the bottom (left) as an image pair. The dark line represents the aligned hollow spaces formed by the middle portions of neighboring N-monomers interacting with RNA. In the further zoomed version at the bottom (right) two N-monomers (shown on one of the duplicate images in red and yellow) are positioned in the middle of the field. They appear as dumbbell shapes, with their termini touching each other. The termini appear bright while the middle portions of the neighboring N-monomers surround a darker area. The dark line running in the middle, across the chain of these hollow spaces between the monomers (drawn in blue), is most likely the place where RNA is associated with the N-protein. This arrangement resembles the models of RNPs of vesicular stomatitis and rabies viruses.[13,14,29]

Large round structures were found abundant in the cryo-EM micrographs of recombinant hantaviral N protein (Fig. 6). These coiled ring-like (sometimes, arch-like) structures are very similar, both

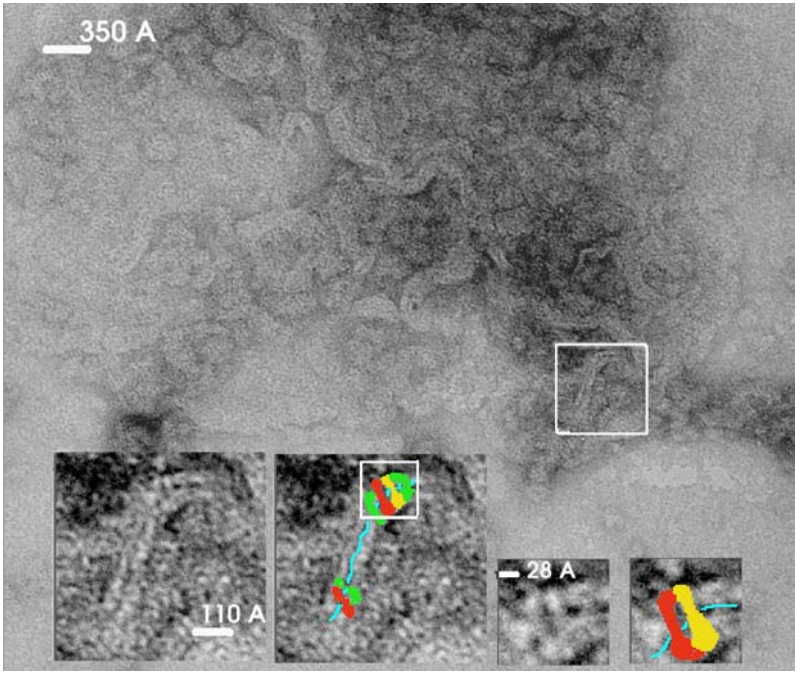


Fig. 5. Fibrous RNP complexes formed by recombinant hantaviral N protein and cellular RNA. The fibrous complexes seen in the background are interpreted as N protein associated with RNA. They are detected in samples negatively stained with uranyl acetate (see Fig. 2). The selected region (framed white) is shown at higher magnification in the two insertions at the bottom left. The N-protein monomers are colored red, green, and yellow, and the possible track of underlying RNA that the N-monomers are covering is shown as a blue line. The white frame in the insertion is shown in detail as a pair of smaller insertions to the right; identified monomers of the N protein are colored red and yellow and the suggested track of RNA is shown in blue. Our interpretation of the arrangement of N-monomers in fibrous complexes is in remarkable agreement with the models presented for the rabies virus nucleocapsid (cf. Figs. 6b and 6d in Schoehn *et al.*, 2001[29]).

in size and shape, to structures formed by recombinant N proteins of influenza[28] and rabies[29] viruses and, most probably, are oligomers of the N protein with cellular RNA wrapped. The rings of this size can accommodate 10 to 12 monomers of the hantaviral N protein; they are probably produced from larger structures by breaking the RNA

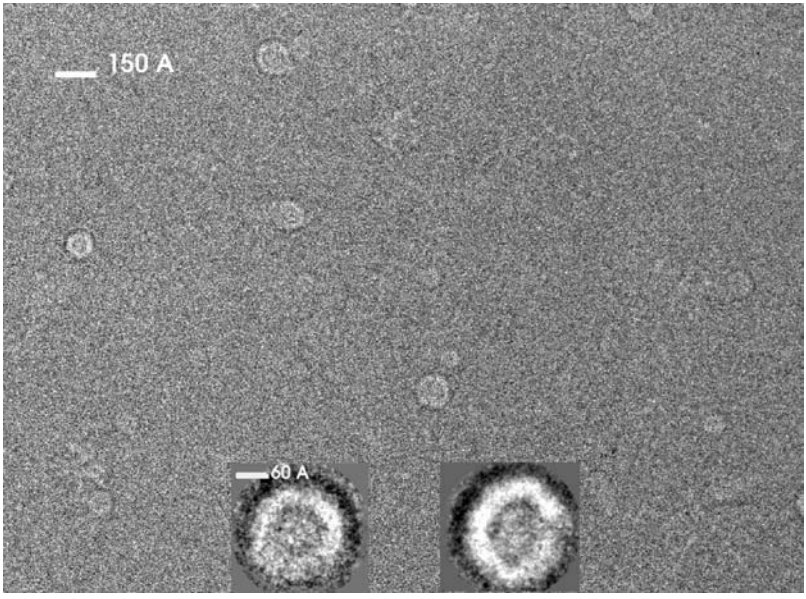


Fig. 6. Cryo-EM preparation of purified recombinant Puumala virus N protein (see Fig. 2). The standard cryo-EM preparative method was used after a 3 μL sample was applied to glow-discharged holey-carbon 300- or 400-mesh-Au grids (Quantifoil) washed from urea (i.e. floated in sequence in drops of 5 mM Tris-HCl, pH 8.0). In the cryo-EM images, rings and arches are visible that suggest higher-order packing of RNP. In the insets, class averages of particles are shown at higher magnification. Reliable 3D reconstructions were difficult to obtain due to heterogeneities in particle sizes. The class averages were calculated using reference-free classification with EMAN.

during preparation of samples. The monomers in these structures appeared to be tightly packed, because no indications of cavities were seen. Tight packaging of the twisted monomers in the N-protein oligomers is in agreement with recent crystallographic data on RNPs of rabies and vesicular stomatitis viruses.[13,14]

Concluding Remarks

A single-particle reconstruction with three-fold axis of rotational symmetry (C3) revealed the existence, in addition to monomers, of an

N-trimer in purified preparations of recombinant hantaviral N protein (produced in the baculovirus-driven system in insect cells). This was not totally unexpected as the hantaviral N protein was previously shown to form trimers.[16,17,19] Our reconstruction of the hantaviral N-monomers showed a striking similarity to structures described in the literature: (1) earlier, low-resolution single-particle reconstructions of the N-protein complexes of rabies virus,[29] and (2) the most recent, high-resolution crystallography data for RNA complexes of rabies and vesicular stomatitis viruses.[13,14] For example, the twisted shape characteristic for hantaviral N-monomers (Fig. 2) is apparent in the tightly formed RNP of rabies virus (see the space-filling model on Fig. 3A in Albertini *et al.*, 2006[13]). A relatively thin link between N-terminal and C-terminal domains in the nucleoproteins of rabies and vesicular stomatitis viruses (see Fig. 1C in Albertini *et al.*, 2006[13] and Fig. 1B in Green *et al.*, 2006[14]) can be easily recognized in the hantaviral N-trimer (Fig. 2C, tilted side view, marked with arrow).

Reconstruction of the N-trimer structure, together with the RNA docking data, led to the hypothesis on the mechanism of RNA chaperoning by the hantaviral N-trimer. The main idea is that the vRNA panhandle is dissociated by the trimer as a result of solving a steric conflict presented by the fitting together of two highly specific 3D structures formed by the interacting parties. Here it is important to stress that single monomers cannot perform the task and it is a specific structure, the N-trimer, which supposedly acts as a chaperone. It remains to be seen whether N-oligomers of other NSRVs can act as RNA chaperones, and if so, what would be the underlying mechanism.

Striking similarities have been observed between RNP complexes formed by the hantaviral N protein with non-specific RNA and corresponding structures described for the following other NSRVs: influenza, vesicular stomatitis and rabies viruses. The ring-like structures formed by 10 to 12 N-monomers associated with RNA were most noticeable. These findings suggest a common mechanism of formation and functioning of viral RNPs. It is also worth mentioning that, in the higher-order RNP-structures, the hantaviral N-trimers are not seen. This could be explained by the absence of virus-specific, panhandle-forming RNA sequences in the pool of encapsidated cellular

RNA, and it points to a possibility that the N-trimer acts merely as an intermediate during the viral RNA encapsidation.

Acknowledgment

We thank Pasi Laurinmäki from Dr Sarah Butcher's Cryo-Electron Microscopy Unit, University of Helsinki, for technical help in collecting the cryo-electron microscopy raw data, and Professor AnttiVaheri, Department of Virology, University of Helsinki, for his support.

References

1. Nichol ST, Beaty BJ, Elliott RM, *et al.* (2005) In C.M. Fauquet, M.A. Mayo, J. Maniloff, *et al* (eds), *Virus taxonomy. VIIIth report of the International Committee on Taxonomy of Viruses,* pp. 695–716. Elsevier Academic Press, Amsterdam.
2. Kukkonen SKJ, Vaheri A, Plyusnin A. (2005) L protein (RNA-dependent RNA polymerase) of hantaviruses (Review). *Arch Virol* **150**: 533–556.
3. Gavrilovskaya IN, Shepley M, Shaw R, *et al.* (1998) Beta3 integrins mediate the cellular entry of hantaviruses that cause respiratory failure. *Proc Natl Acad Sci USA* **95**: 7074–7079.
4. Jin M, Park J, Lee S, *et al.* (2002) Hantaan virus enters cells by clathrin-dependent receptor-mediated endocytosis. *Virology* **294**: 60–69.
5. Garcin D, Lezzi M, Dobbs M, *et al.* (1995) The 5′ ends of Hantaan virus *(Bunyaviridae)* RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis. *J Virol* **69**: 5754–5762.
6. Jonsson CB, Schmaljohn CS. (2001) Replication of Hantaviruses. In C.S. Schmaljohn & S.T. Nichol (eds), *Hantaviruses,* pp. 15–32. Springer-Verlag, Berlin Heidelberg.
7. Kariwa H, Tanabe H, Mizutani T, *et al.* (2003) Synthesis of Seoul virus RNA and structural proteins in cultured cells. *Arch Virol* **148**: 1671–1685.
8. Kaukinen P, Vaheri A, Plyusnin A. (2005) Hantavirus nucleocapsid protein: a multifunctional molecule with both housekeeping and ambassadorial duties (Review). *Arch Virol* **150**: 1693–1713.
9. Xu X, Severson W, Villegas N, *et al.* (2002) The RNA binding domain of the hantaan virus N protein maps to a central, conserved region. *J Virol* **76**: 3301–3308.

10. Severson WE, Xu X, Jonsson CB. (2001) cis-Acting signals in encapsidation of Hantaan virus S-segment viral genomic RNA by its N protein. *J Virol* **75**: 2646–2652.
11. Mir MA, Panganiban AT. (2004) Trimeric hantavirus nucleocapsid protein binds specifically to the viral RNA panhandle. *J Virol* **78**: 8281–8288.
12. Mir MA, Panganiban AT. (2005) The hantavirus nucleocapsid protein recognizes specific features of the viral RNA panhandle and is altered in conformation upon RNA binding. *J Virol* **79**: 1824–1835.
13. Albertini AAV, Wernimont AK, Muziol T, *et al.* Crystal structure of the rabies virus nucleoprotein-RNA complex. *Sciencexpress*, 15 June 2006; 1125280.
14. Green T, Zhang X, Wertz GW, Lung M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Sciencexpress*, 15 June 2006; 1126953.
15. Mir MA, Panganiban AT. (2006) The bunyavirus nucleocapsid protein is an RNA chaperone: possible roles in viral RNA panhandle formation and genome replication. *RNA* **12**: 272–282.
16. Kaukinen P, Koistinen V, Vapalahti O, *et al.* (2001) Interaction between molecules of hantavirus nucleocapsid protein. *J Gen Virol* **82**: 1845–1853.
17. Alfadhli A, Love Z, Arvidson B, *et al.* (2001) Hantavirus nucleocapsid protein oligomerization. *J Virol* **75**: 2019–2023.
18. Kaukinen P, Vaheri A, Plyusnin A. (2003) Mapping of the regions involved in homotypic interactions of Tula hantavirus N protein. *J Virol* **77**: 10910–10916.
19. Kaukinen P, Kumar V, Tulimäki K, *et al.* (2004) Oligomerization of Hantavirus N protein: C-terminal alpha-helices interact to form a shared hydrophobic space. *J Virol* **78**: 13669–13677.
20. Yoshimatsu K, Lee BH, Araki K, *et al.* (2003) The multimerization of hantavirus nucleocapsid protein depends on type-specific epitopes. *J Virol* **77**: 943–952.
21. Alminaite A, Halttunen V, Kumar V, *et al.* (2006) Oligomerization of hantavirus nucleocapsid protein: analysis of the N-terminal coiled-coil domains. *J Virol* **80**: 9073–9081.
22. Alfadhli A, Steel E, Finlay L, *et al.* (2002) Hantavirus nucleocapsid protein coiled-coil domains. *J Biol Chem* **277**: 27103–27108.
23. Flick K, Hooper JW, Schmaljohn CS, *et al.* (2003) Rescue of Hantaan virus minigenomes. *Virology* **306**: 219–224.

24. Kukkonen SKJ, Vaheri A, Plyusnin A. (2004) Tula hantavirus L protein is a 250 kDa perinuclear membrane-associated protein. *J Gen Virol* **85**: 1181–1189.
25. Pettersson RF, Melin L. (1996) Synthesis, assembly and intracellular transport of *Bunyaviridae* membrane proteins. In R.M. Elliott (ed), *The Bunyaviridae*, pp.159–188. Plenum Press, New York.
26. Vapalahti O, Lundkvist Å, Kallio-Kokko H, *et al.* (1996) Antigenic properties and diagnostic potential of Puumala virus nucleocapsid protein expressed in insect cells. *J Clin Microbiol* **34**: 119–125.
27. Kumar V, Heikkonen J, Engelhardt P, Kaski K. (2003) Robust filtering and particle picking in micrograph images towards 3D reconstruction of purified proteins with cryo-electron microscopy. *J Struct Biol* **145**: 41–51.
28. Martin-Benito J, Area E, Ortega J, *et al.* (2001) Three-dimensional reconstruction of a recombinant influenza virus ribonucleoprotein particle. *EMBO Reports* **21**: 313–317.
29. Schoehn G, Iseni F, Mavrakis M, *et al.* (2001) Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J Virol* **75**: 490–498.
30. Stewart L, Matthew RR, Qiu X, *et al.* (1998) A model for the mechanism of human topoisomerase. *Science* **279**: 1534–1542 .
31. Martin SL, Branciforte D, Keller D, Bain DL. (2003) Trimeric structure for an essential protein in L1 retrotransposition. *Proc Natl Acad Sci USA* **100**: 13815–13820.

Chapter 22

Astrovirus Replication: An Overview

Susana Guix[†], *Albert Bosch*[*,†] *and Rosa M. Pintó*[†]

Human astroviruses are important pathogens that cause gastroenteritis worldwide. Significant progress has recently been made regarding the characterization of the RNA replication process, the apoptotic response induced in virus-infected cells, and the formation of virus-like particles. First, a relationship between astrovirus RNA replication sites and the endoplasmic reticulum-derived membranes has been suggested. In addition, a direct relationship between nonstructural proteins and the induction of apoptosis has been observed, and it has been demonstrated that apoptotic host cell death seems necessary for maturation of astrovirus particles. Finally, it has been predicted that the VP34 capsid protein contains the RNA binding domain at its N-terminus, responsible for packaging the viral genome, as well as an 8 β-barrel domain that may likely constitute the building subunit to form the T = 3 icosahedral capsid. The C-terminal half of the capsid polyprotein — highly variable between different astroviruses — is thought to form the receptor-interaction domain.

Background for Human Astroviruses

Infections of the gastrointestinal tract cause approximately two billion cases of diarrhea in children per year worldwide, with the majority of

*Corresponding author: Albert Bosch, Department of Microbiology, University of Barcelona, Avda Diagonal 645, 08028 Barcelona, Spain, E-mail: abosch@ub.edu.
†Enteric Virus Group, Department of Microbiology, University of Barcelona.

deaths occurring in developing countries.[1] Diarrhea is the third leading cause of mortality worldwide. Every year, an estimated three million pediatric deaths result from viral gastroenteritis and dehydration.[2] Along with rotavirus, calicivirus, and enteric adenovirus, human astroviruses (HAstV) are one of the most important causes of viral pediatric acute gastroenteritis.[3–5] HAstV were first identified 30 years ago by Appleton and Higgins[6] as the cause of a gastroenteritis outbreak in a maternity ward. Shortly thereafter, Madeley and Cosgrove[7] proposed the name "astrovirus" after observing the virus' star-like morphology by electron microscopy (EM) in stool samples from a pediatric diarrhea outbreak. Despite the years that have elapsed since HAstV were discovered, many aspects of their molecular biology and pathogenesis remain still to be elucidated.

Human astroviruses are non-enveloped icosahedral viruses that belong to the family *Astroviridae* with a plus-sense, single-stranded RNA genome. According to the International Committee on Taxonomy of Viruses, the *Astroviridae* family is divided into two genera: *Mamastrovirus*, which includes astroviruses that infect mammals and primarily causes gastroenteritis, and *Avastrovirus*, which includes astroviruses that infect avian species and may cause intestinal as well as extra-intestinal illness. Presently, HAstV are divided into eight serotypes (HAstV-1 to HAstV-8), with HAstV-1 being the most prevalent worldwide. HAstV-1 to 7 were initially identified according to the reactivity of the capsid proteins with reference type-specific rabbit antibodies, while HAstV-8 was fully characterized some years later.[8–10] Over the last few years, the number of astrovirus sequences available via Genbank has rapidly increased and the complete genome for five HAstV isolates — including serotypes 1, 2, 3, and 8 — has been sequenced.[9,11–14] Complete capsid sequences are available for HAstV-1 through 6 and HAstV-8, as well as for feline, porcine, ovine, mink, and turkey astroviruses and for the avian nephritis virus. HAstV can be isolated and propagated in human intestinal CaCo-2 continuous cell line in the presence of trypsin, which is involved in the capsid protein maturation.[15] An infectious full-length cDNA clone is available for HAstV-1.[16] BHK-21 cells can support HAstV-1 RNA

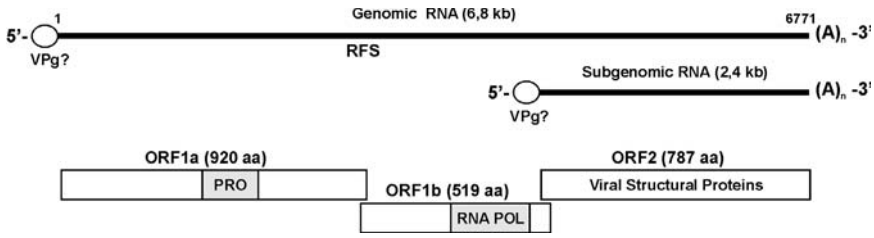


Fig. 1. Genomic organization of HAstV. Genomic and subgenomic RNA molecules are indicated in the upper panel, and ORFs with their main functional motifs are shown in rectangles. Ribosomal frame-shifting signal (RFS) between ORF1a and ORF1b is indicated. The presence of a VPg protein at the 5' end of RNA has not been confirmed. Nucleotide and amino acid positions are numbered according to HAstV-1 Oxford reference strain (accession no. L23513).

replication when transfected with *in vitro* transcribed RNA, and virions produced in these cells can infect CaCo-2 cell monolayers in trypsin-containing media.

Virions of HAstV contain a positive sense single-stranded 6.8 kb RNA genome, which is polyadenylated at the 3'-end and includes a 5'-untranslated region (UTR), three overlapping open reading frames (ORFs), and a 3'-UTR (Fig. 1). Although suggested, the presence of a covalently linked VPg protein at the 5'-end of the genome has not yet been biologically demonstrated.[17] The fact that *in vitro* transcribed capped RNA is infectious does not rule out the possibility that the 5' end of genomic RNA found within particles could be linked to a VPg protein, and similar observations have been made with feline calicivirus.[18] ORF1a and ORF1b are linked by a ribosomal frame-shifting signal (RFS) and code for the nonstructural proteins, including a serine protease and an RNA-dependent RNA polymerase, while ORF2 encodes the capsid precursor.[11]

Although it is considered characteristic of positive-stranded RNA viruses with genomes larger that 6 kb to encode a helicase domain, there was a widespread belief for many years that astroviruses did not code for a helicase protein. However, two putative helicase conserved

motifs upstream from the protease coding region, which show sequence similarity to pestivirus RNA helicase sequences, have recently been mapped by computational tools.[17]

Upon infection, two large nonstructural proteins are translated from the genomic RNA: nsP1a and nsP1a1b. These proteins mature by proteolysis involving both the viral protease and cellular proteases, and give rise to the viral proteins implicated in the transcription of a full-length negative-strand RNA molecule (antigenomic RNA). It is believed that this RNA molecule serves as a template for the transcription of both the new full-length 6.8 kb-genomic and the ORF2-containing 2.4 kb-subgenomic RNA molecule.[19] Translation of this subgenomic RNA is used to produce large amounts of structural proteins for the efficient assembly of the progeny viruses.[15] Both astrovirus genome organization and replication/translation strategies have frequently been compared to those of alphaviruses and caliciviruses due to their similarities.

This review will focus on recent advances in the understanding of the molecular biology, structure, and replication cycle of HAstV. Current efforts are being undertaken to complete the characterization of their genomic organization, to study virion assembly and morphogenesis, and to understand the relationships between the virus and the host immune response. Major developments on these issues are expected in the near future.

Molecular Virology and Replication Cycle: New Insights

Structural Proteins

Considerable progress has been made in understanding the basics of HAstV genomic organization; however, capsid protein processing and assembly, as well as genome encapsidation, are not yet clear. Translation of the structural polyprotein from ORF2 (782 to 794 amino acids, depending on serotype) most likely occurs from the subgenomic RNA. The presence of a conserved RNA sequence upstream from the ORF2 initiation codon with a possible role in the regulation of the

capsid gene expression has been reported.[20] This region has also been identified as a hotspot for RNA recombination.[21]

Proteolytic Processing

As illustrated in Fig. 2, the polyprotein precursor would be cleaved intracellularly by cellular proteases and extracellularly by trypsin to produce three mature structural proteins of 32–34 kDa (VP34), 29–31 kDa (VP29), and 24–26 kDa (VP26), with molecular weights and trypsin cleavage sites varying between different serotypes.[8,15,22–24] In studies based on HAstV-2, it was found that the 87 kDa polyprotein could be assembled into particles and then cleaved into proteins of 32, 29 and 26 kDa.[24,25] Bass and Qiu suggested that assembly of HAstV-1 particles requires an intracellular cleavage of the first 70 N-terminal amino acids of the 87 kDa precursor protein.[22] The resultant 79 kDa protein would assemble into the viral capsid within infected cells. In the absence of trypsin, these particles are virtually noninfectious, but they can become infectious when trypsin cleaves the 79 kDa protein into the final structural proteins VP34, VP29 and VP26. This intracellular cleavage at the N-terminus was not observed in other studies with HAstV-1[26] and HAstV-8[8] infected cells, but it has recently been demonstrated for HAstV-1 that the 70 amino acids at the N-terminus of the polyprotein are not essential to assemble into virus-like particles.[27]

An extensive study using HAstV-8 demonstrated that the 90 kDa polyprotein (VP90) is first cleaved intracellularly at its C-terminus, giving rise to a 70 kDa protein (VP70) that is mainly found into purified particles.[8] The VP70-containing virus is minimally infectious and requires trypsin to enhance its infectivity. Following trypsin activation, VP70 is processed into the three predominant smaller mature products. In a later study, the same authors observed that indeed the capsid precursor VP90 is able to assemble into virions, as it had been shown to do in other studies for HAstV-1[26] and HAstV-2.[24,25] However, these resulting particles were relatively unstable and would disassemble during cesium chloride purification, indicating that the cleavage that yields VP70 structurally stabilizes the viral particles.[23]

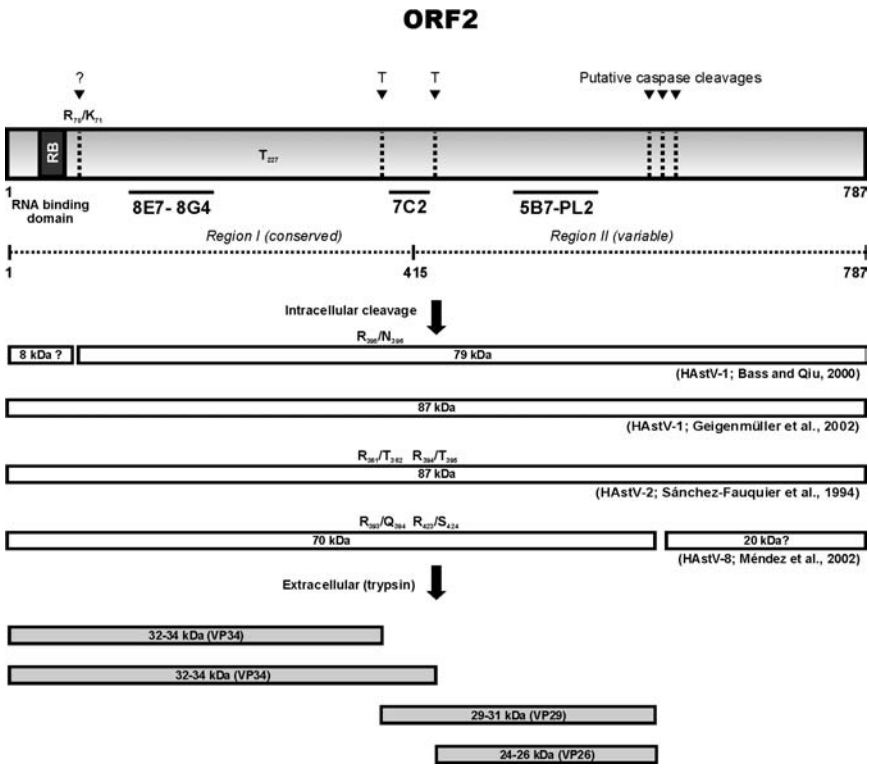


Fig. 2. Schematic depiction of the ORF2 polyprotein of HAstV, and proteolytic processing pathways proposed by different authors using different serotypes (Geigenmüller *et al.*, 2002).[8,22,24] Putative RNA binding domain (RB) at the N-terminus of the polyprotein is indicated. Inverted arrowheads denote putative cleavage sites dependent upon caspases, other unknown cellular proteases (?), and trypsin (T). Underlined regions correspond to the main viral epitopes identified by different authors,[24,37,38] being the epitopes defined by the mAb 8E7 and 8G4 shared by all serotypes. The dotted line indicates the two domains of the protein sequence (amino acids 1–415, well conserved; amino acids 416–end, variable). White boxes represent intermediate products found within infected cells, while grey boxes represent the final products present in infectious particles. See text for details. Amino acid positions are numbered according to HAstV-1 Oxford reference strain (accession no. L23513).

Capsid Protein Maturation and Apoptosis

The fact that HAstV infection induces apoptosis on CaCo-2 cells and that apoptotic death of host cells seems necessary for efficient human astrovirus replication and particle maturation has been demonstrated by different authors.[23,28] Guix *et al.* observed that in the presence of a caspase 8-specific inhibitor, there was a significant reduction in the infectivity of the virus progeny, while the titer of physical virions in cell supernatants estimated by ELISA and RT-PCR was not affected.[28] Méndez *et al.* demonstrated that VP90-VP70 processing is mediated by caspases, and also observed that the release of infectious viruses to the cell supernatant was reduced in the presence of a pancaspase inhibitor and increased by the TNF-related apoptosis-inducing ligand (TRAIL).[23] However, inhibition of apoptosis also resulted in a decrease in the total amount of viral protein detected in cell supernatants, suggesting that processing of VP90 and/or the apoptotic cell response may also affect viral release.[23] Differences between the conditions of infections and the techniques used to measure the amount of released viruses could partially explain these discrepancies. For several viruses it has been observed that some of their proteins, both structural and nonstructural, undergo caspase-mediated proteolysis within the host cell,[29–34] and in some cases such as human influenza virus or Aleutian mink disease parvovirus, these cleavages are relevant for virus replication and morphogenesis.[30,34]

Structural Studies

Very little is known about the structural features of the HAstV capsid, and most of our current knowledge comes from structural predictions and comparison to other related viral families (for a review see Krishna, 2005).[35] Based on the homology between different astroviruses, the ORF2 sequence is divided into two regions. At the N-terminus, region I spans amino acid residues 1 to 415 and is highly conserved among all members of the *Astroviridae* family, while region II at the C-terminus is extremely variable among different strains

(Fig. 2). Wang *et al.* further subdivided region II into three regions with different degrees of similarity.[36] Consistent with the capsid polyprotein domain organization, studies using antibodies revealed that VP34 would express conserved epitopes shared by all serotypes, while VP29 and VP26 contained the serotype-specific neutralizing epitopes.[24,37,38] Thus, it is reasonable to hypothesize that the hypervariable C-terminal region would be located on the surface of the viral particle contributing to the strain-specific tropism of the virus, while the conserved region would constitute the "assembly domain," a building block for capsid formation and encapsidation of the viral genome.[27,35] Interestingly, during the construction of the HAstV-1 infectious clone it was found that amino acid residue $Thr_{227,}$ within region I of the capsid polyprotein, was essential for proper assembly of viral capsids.[19]

Finally, within the first 70 residues of the ORF2 polyprotein, region I includes a well-conserved domain rich in basic amino acids, which has been associated with the viral RNA packaging process due to its potential RNA-binding properties and its similarities to other well-documented icosahedral RNA viruses, such as Sindbis virus and some plant viruses.[39–41] Consistent with this observation, it has been shown that the first 70 amino acid residues of HAstV-1 are dispensable for virus-like particle formation using a recombinant baculovirus expression system.[27] Jonassen *et al.* also related this arginine-rich region to gene expression regulation processes in other viruses such as coronaviruses, papillomavirus and baculovirus.[59]

Detailed ultrastructure studies of trypsin-treated HAstV-2 virions revealed icosahedral particles with an array of spikes protruding from the surface of the virion.[42] A low-resolution cryoelectron microscopy image for trypsin-treated HAstV-1 particles showed a solid icosahedral capsid shell 330 Å in diameter and decorated with 30 dimeric spikes extending 50 Å from the surface.[4]

Using the three-dimensional position-specific scoring matrix (3D-PSSM)[43] to identify recognizable protein-folding motifs within the astrovirus capsid proteins, meaningful structural predictions are currently providing insights into different functional aspects of the viral

capsid biology. This approach has recently been used by two different research groups, leading to similar conclusions.[27,35] Consistently, all significant matches for VP34 sequence from different HAstV serotypes and other animal astroviruses correspond to coat proteins from simple, icosahedrally symmetric viruses with jelly-roll β-barrel subunits, such as Theiler's murine encephalomyelitis virus, bean pod mottle virus, carnation mottle virus, and tomato bushy stunt virus. Structural alignments of predicted structural homology comparisons between a fragment of VP34 sequence and coat proteins from these viruses are shown in Fig. 3. Interestingly, the putative RNA-binding domain present at the N-terminus of VP34 was never included in positive alignments, confirming a different biological function for these two domains of the VP34 protein.

The evaluation of protein-folding prediction for the variable domain (amino acid 416 to the end of ORF2) showed significant structural homology to virus-related proteins and receptors, as well as to some non-viral receptor-ligand interactions.[35] These results support the idea that this capsid protein domain is responsible for the receptor-interaction process, which dictates cell tropism and may vary enormously between different astrovirus strains.[35] Of interest, most significant alignments did not include the C-terminal region of ORF2, suggesting that this sequence would not be included in the infectious particle, as some of the proposed models of morphogenesis suggest.[8]

These prediction results would indicate that capsid proteins from all astroviruses fold in a manner similar to that of well-studied ssRNA icosahedral viruses with T = 3 symmetry, and that the predicted β-barrel domain found in VP34 protein could be the building block for capsid assembly. One common feature of ssRNA viruses that show T = 3 symmetry is that their capsids are made of a single structural building block. Assuming that 180 copies of the VP34 protein are required to give a particle of 33 nm with T = 3 symmetry and that the spikes are formed by VP26 and VP29 proteins would be consistent with the structural prediction that suggests that VP34 is a structural homolog to known icosahedral coat proteins and that the variable domain (VP26/VP29) is involved in receptor binding.

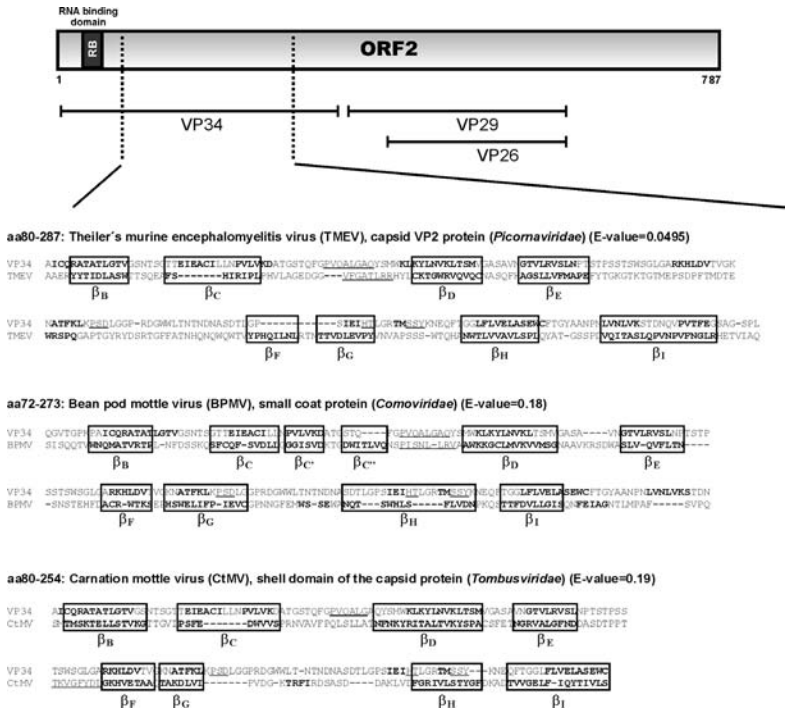

RNA binding domain
RB
ORF2
1
787
VP34
VP29
VP26
aa80-287: Theiler´s murine encephalomyelitis virus (TMEV), capsid VP2 protein (Picornaviridae) (E-value=0.0495)
aa72-273: Bean pod mottle virus (BPMV), small coat protein (Comoviridae) (E-value=0.18)
aa80-254: Carnation mottle virus (CtMV), shell domain of the capsid protein (Tombusviridae) (E-value=0.19)

Fig. 3. Schematic representation of the predicted eight β-barrel domain structure within the VP34 capsid protein of HAstV and three structure-based sequence alignments of HAstV sequence with capsid proteins from other viruses: VP2 coat protein of Theiler's murine encephalomyelitis virus (Protein Data Bank 1TMF), small subunit of Bean pod mottle virus coat protein (Protein Data Bank 1PGL), and coat protein (shell domain) of Carnation mottle virus (Protein Data Bank 1OPO). E-value is the measure of confidence in prediction of the 3D-PSSM alignment. Beta-barrel domains are indicated by boxes, beta-strands are indicated in bold, and underlined sequences correspond to alpha-helices. RB denotes the putative RNA binding domain. Amino acid (aa) positions refer to HAstV-1 Oxford reference strain (accession no. L23513).

HAstV Virus-like Particles (VLPs)

Recently, it has been demonstrated that expression of the complete ORF2 of HAstV serotypes 1 and 2 leads to the formation of VLPs using the baculovirus and vaccinia virus recombinant expression systems, respectively.[25,27] Using the baculovirus system, VLPs can be

obtained either with the expression of a complete ORF2, a truncated form of ORF2 lacking the first 70 amino acids, or after replacing these 70 amino acids with the green fluorescent protein (GFP). Recombinant HAstV VLPs will provide a virtually unlimited supply of highly purified viral capsids with many immediate applications, such as the characterization of antigenic properties of the mature capsid proteins and epitope mapping studies. Additionally, VLPs may also be used to determine the HAstV atomic structure using X-ray crystallographic techniques. Another potential application of the expressed recombinant VLPs is to study cell-virus interactions in binding assays, which may allow the identification of the cellular receptor involved in viral infection.

Interestingly, during the formation of HAstV VLPs in insect cells, a smaller second type of structure consisting of 16 nm ring-like units was observed, mostly after disassembling the 38 nm VLPs through the addition of EDTA.[27] As previously described for Norwalk virus as well as for other T = 3 RNA plant viruses, these 16 nm VLPs could correspond to structures that are formed when 60 units of the capsid protein assemble into a structure with T = 1 symmetry. Depending on factors such as trypsin digestion, the presence and size of the RNA genome, the presence of divalent cations, the ionic strength and the pH, capsid proteins of these viruses are able to self-assemble *in vitro* into T = 3 or smaller T = 1 structures.[44,45]

Nonstructural Proteins

The nonstructural proteins of the virus are translated from the genomic viral RNA as two polyproteins, one of which contains only ORF1a (nsP1a: 101 kDa) and the other that includes ORF1a/1b (nsP1a1b: 160 kDa) and is translated via a –1 ribosomal frame-shifting event between ORF1a and ORF1b.[11] Both proteins are proteolytically processed, giving rise to a variety of proteins. While nsP1b (59 kDa) corresponds to the viral RNA-dependent RNA polymerase — which can be aligned with other polymerases of the Koonin's supergroup I together with picornaviruses, caliciviruses and certain plant viruses — little is known about the role of most of the mature nsP1a products (Fig. 4). The protease motif has features consistent with chymotrypsin-like

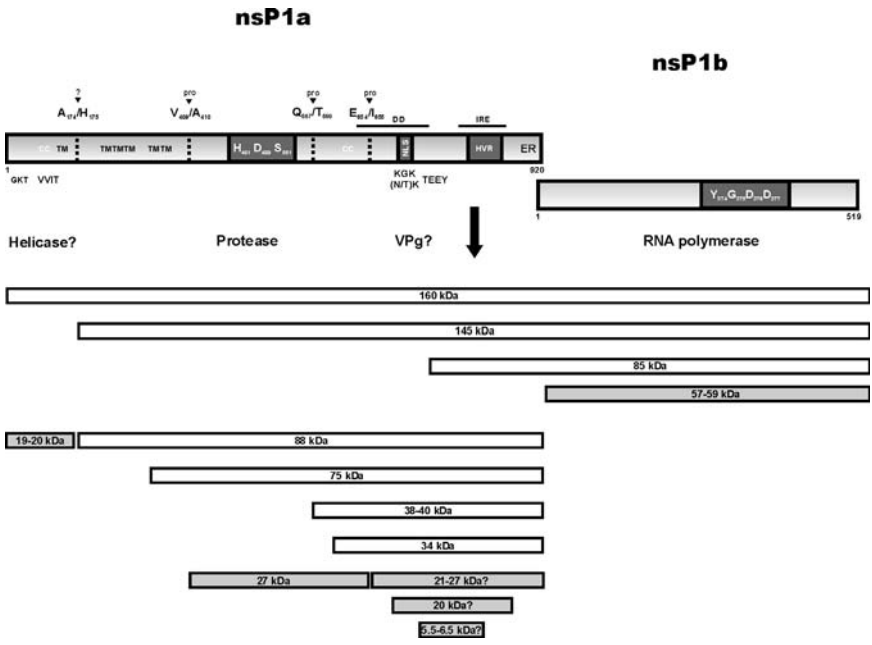


Fig. 4. Schematic depiction of HAstV nonstructural polyproteins nsP1a and nsP1b, and proposed proteolytic processing pathway. Predicted transmembrane helices (TM), nuclear localization signal (NLS), immunoreactive epitope (IRE), coiled-coil structures (CC), predicted death domain (DD), hyper-variable region (HVR), KKXX-like endoplasmic reticulum retention signal (ER), protease motif, RNA polymerase motif, and putative RNA helicase and VPg motifs are shown. Inverted arrowheads denote putative cleavage sites dependent on unknown cellular proteases (?), and the viral protease (pro). White boxes represent intermediate products, while grey boxes represent the final products. See text for details. Amino acid positions are numbered according to HAstV-1 Oxford reference strain (accession no. L23513).

proteases of other plus-stranded RNA viruses, including rabbit hemorrhagic disease virus and feline calicivirus,[11] with the substitution of a serine for a cysteine at the third catalytic amino acid residue (His_{461}, Asp_{489}, and Ser_{551}). Some of the amino acid residues involved in substrate binding involve Thr_{546} and His_{566}.[11,46]

Several transmembrane helices (TM) have been identified at the N-terminus of nsP1a. Although the exact number varies among different studies,[11,20] it is believed that these TMs are responsible for

the anchorage of the RNA replication complex to intracellular membranes. Based on the software predictions on the orientation of the TM helices, it seems likely that the N-terminal region of the helices in nsP1a is on the cytosolic side of a membrane. Sequence analysis has also predicted a KKXX-like endoplasmic retention (ER) signal that suggests an association between the viral RNA replication process and ER-derived intracellular membranes.[47] In addition, computer prediction led to the identification of a bipartite nuclear localization signal (NLS).[11] Although some nsP1a-derived products have been shown to accumulate in the nuclei of infected cells,[48] the significance of a nuclear involvement in an RNA virus lifecycle is yet to be understood. Recently, two common motifs for astrovirus and pestivirus RNA helicase (GKT and VVIT) have been proposed upstream from the protease coding region. These motifs would represent motifs I and II of the seven conserved motifs of SF2 helicases and would be functionally important in ATP binding.[17]

The presence of an immunoreactive epitope (IRE) close to the C-terminus of nsP1a was partially characterized some years ago, indicating the high immunogenicity of this region.[49] In addition, a hypervariable region (HVR) has been identified close to this epitope by different authors following genetic characterization of HAstV-1, HAstV-3, HAstV-4, and HAstV-8.[9,13,47,50] Variability within this HVR has been associated with different viral RNA replication and growth properties, as well as with different virus RNA levels in feces from children with gastroenteritis, suggesting a relationship between certain genotypes and some viral properties related to its pathogenic phenotype.[51] Variability within this region consists mainly of high rates of nucleotide and amino acid substitutions, as well as many insertions and deletions that retain the reading frame. Interestingly, a 15-amino acid deletion identified some years ago was related to adaptation of HAstV to certain cell lines,[50] and recently this region has been associated with a distinctive RNA replication pattern. Although the molecular mechanisms that regulate the efficient minus and plus RNA strands, including the synthesis of a subgenomic RNA, remain unclear, mutagenesis studies have demonstrated that genetic variability of the C-terminal of nsP1a affects the virus RNA replication phenotype.[51] Additionally, using antibodies against the HVR, it has been shown

that the C-terminal protein of nsP1a colocalizes with the endoplasmic reticulum and viral RNA in HAstV-4 CaCo-2 infected cells, suggesting the involvement of this protein in the RNA replication process in endoplasmic reticulum-derived intracellular membranes.[47]

Computer analysis of nsP1a has recently revealed the presence of two coiled-coil regions (CC) common to all known human astroviruses, which are hypothesized to be involved in the formation of protein oligomers.[20] The absence of a methyltransferase-encoding region and the similarity of the HAstV RNA-dependent RNA polymerase motif with primarily polymerases of VPg-containing viruses raise the possibility that the HAstV genome, as well as the subgenomic RNA, may be linked to a VPg protein in its 5'-end. A convincing VPg domain has not been identified, but it may be located upstream from the protease motif, where it has been postulated that Ser_{420} could link the VPg to viral RNA.[11] However, based on sequence comparison between human and animal astroviruses with some members of the *Caliciviridae*, *Picornaviridae* and *Potyviridae* families, other authors have also suggested that amino acid Tyr_{693} at the conserved TEEEY-like motif could also display a VPg function.[17,20] Indeed, one of the conserved amino acid motifs characteristic of VPg of caliciviruses [KGK(N/T)K][52] can also be identified upstream from this Tyr_{693} residue at a similar distance as it is found in the calicivirus genome.

Finally, it has recently been reported that products derived from nsP1a lead to apoptosis of the host cell, resulting in efficient virus replication and particle release.[23,28] The analysis of the secondary structure of the whole astrovirus protein sequence revealed the presence of an optimal six α-helix structure of 95 amino acids close to the C-terminus of nsP1a, which displays homology to the death domain superfamily.[28] The presence of a putative death domain in nsP1a suggests a direct link between this polyprotein and the apoptotic pathway.

Proteolytic Processing

The proteolytic processing of the HAstV nsP1a polyprotein (920–935 amino acids, depending on serotype) has been only partially characterized, and some of the reported data are still conflicting.[40,46,48,53,54]

Data obtained with *in vitro* studies with HAstV-1 suggest that it is cleaved at Gln_{567}/Thr_{568} in a process that is dependent on the viral 3C-like serine protease, resulting in an N-terminal p64 fragment that includes the protease motif and a C-terminal p38 product that includes the NLS and the IRE.[46] However, using a similar approach with a different serotype, Gibson *et al.* could not identify any proteolytic activity,[53] suggesting that the viral protease may require some cellular factors.

A later work performed on transfected BHK-21 cells suggested a proteolytic cleavage site close to amino acid 170 of nsP1a, which would generate an N-terminal product of approximately 20 kDa, and two cleavages sites around amino acid 410 and 655, which would result in a 27 kDa protein containing the protease motif. The viral protease would be responsible for only the two latter cleavages.[40]

Performing pulse chase experiments on HAstV-8 infected CaCo-2 cells, Méndez *et al.* identified polypeptides of 88 and 75 kDa, and polypeptides of 145 and 85 kDa as intermediate products of nsP1a and nsP1a1b proteolytic processing, respectively.[54] As final products, they reported two proteins of 19 and 20 kDa (N-terminal product), two proteins of approximately 27 kDa (protease), and a protein of 57 kDa (RNA polymerase). Results from kinetic experiments were consistent with the idea that a cotranslational processing at the N-terminus of nsP1a and nsP1a1b polyproteins occurs during infection in a process that is independent of the viral protease activity. Based on sequence homology with bovine viral diarrhea virus, the authors suggested that cleavage would occur after the conserved sequence GGYA, between amino acid residues 171 and 174 of nsP1a. Using transient expression experiments on BHK-21 cells, the authors could detect an additional protein of 20 kDa that would correspond to the C-terminal product of nsP1a. Willcocks *et al.* detected products of 75, 34, 20, 6.5, and 5.5 kDa in HAstV-1 infected CaCo-2 cell extracts using antibodies against the 30% C-terminal end of nsP1a, as well as a protein of 59 kDa using antibodies to nsP1b.[48] Using computer predictions, after mapping the VPg domain downstream the protease domain, Al-Mutairy *et al.* suggested potential cleavage sites for the VPg protein.[17] The authors hypothesized that the N-terminal cleavage

site for astrovirus VPg would be located immediately upstream from the KGK(N/T)K motif, and that its C-terminal cleavage site might be found between 92–143 amino acid residues downstream from the N-terminal cleavage site. However, the actual occurrence of a VPg protein within this region, as well as its boundaries, is yet to be experimentally confirmed.

Finally, immunoprecipitation studies on HAstV-4 infected CaCo-2 cells with an antibody against a synthetic peptide corresponding to the HVR sequence close to the C-terminus of nsP1a led to the detection of five proteins in the range of 21–27 kDa, and at least one of them could be post-translationally modified by phosphorylation on its serine and/or threonine residues.[47] In addition, after finding proteins larger than expected in nuclear extracts from recombinant baculovirus-infected cells expressing amino acids 643–940 of nsP1a, Willcocks *et al.* also hypothesized post-translational modifications.[48] Since many proteins destined for the nucleus are heavily glycosylated, the authors suggested that nsP1a proteins are modified by glycosylation.

Structural Computational Predictions

Although it is thought that all nonstructural proteins would form the RNA replication complex and that the transmembrane helices present on nsP1a would anchor the complex to intracellular membranes, the functional properties of most of the individual proteins of HAstV have been poorly characterized. Results obtained after making functional predictions using the 3D-PSSM server for each of the different HAstV uncharacterized non-structural proteins are shown in Table 1. The previously identified sequence with homology to a death domain structure is also included.[28] All matches found, except the protease and the polymerase motifs, displayed just a moderate degree of certainty, and the only region that did not display any significant match corresponded to the N-terminal product, which is rich in transmembrane helices. Amino acids 191–353 showed structural homology to the DNA-protecting protein under starved conditions (DPS protein), a ferritin homolog that unspecifically binds and protects DNA[55]; amino acids 596–739 matched a sarcoplasmic calcium-binding

Table 1. Top Protein Fold Matches for HAstV Nonstructural Proteins using 3D-PSSM Web Server

HAstV Region[a]	Result	Superfamily	Template Length (aa)	Sequence Identity (%)	E-value
nsP1a					
191–353	Dodecameric ferritin homolog DPS	DNA protecting protein during starvation (DPS); Ferritin-like	159	10	0.562
370–655	DegP (HtrA)	Heat shock protease HtrA	299	13	2.68e-10
596–739	Sarcoplasmic calcium-binding protein	EF-hand, calmoduline-like	174	9	0.686
607–737	Human Fas	Death domain	127	22	8.52
718–897	Tumor suppressor bin1 (amphiphysin II)	Myc box dependent interacting protein 1; endocytosis/exocytosis	213	18	1.71
nsP1b					
44–515	Rabbit hemorrhagic disease virus RNA-2 dependent RNA polymerase	RNA-directed RNA polymerase	493	13	0.149e-3

[a]Amino acid (aa) positions are numbered according to HAstV-1 Oxford reference strain (accession no. L23513).

protein, a protein with an EF-hand calmoduline motif; and amino acids 718–897 displayed homology to the tumor suppressor bin1 (amphiphysin II), which belongs to a family of adapter proteins that have been implicated in a variety of different cellular processes, including endocytosis, actin cytoskeletal organization, transcription, and stress responses.[56] Although the sequence corresponding to the HAstV protease showed structural relationship to the protease motif of the heat shock protein DegP (Htra) of *Escherichia coli*, the NCBI conserved domain search identified homology to the Equine arterivirus serine endopeptidase S32 (pfam 05579). Although not contained within the significant alignment, both proteases contain a C-terminal extension domain that may play a role in mediating protein-protein interactions, which may be a novel way of regulating its proteolytic activity.[57,58]

Using similar approaches, an extensive proteomic computational analysis was performed for HAstV-4 nsP1a C-terminal product.[47] This analysis indicated that the proportion of residues that were not assigned to a particular secondary structure was higher within this region of nsP1a than in other regions of the genome, revealing the high degree of structural flexibility that this region may have. Accordingly, analysis of genetic variability revealed that this region was the most variable region of the whole nonstructural coding region, showing a high degree of tolerance to insertions and deletions. The authors also suggested a higher occurrence of post-translational modification motifs, such as phosphorylation and O-glycosylation. Although sequence homology searches only displayed proteins with a moderate degree of similarity, the most significant results involved nonstructural proteins from single-stranded RNA viral families involved in the formation of the RNA replicase complexes and/or in the regulation of both viral and cellular transcription processes, and at the superfamily level, the protein was assigned to the "winged-helix" DNA-binding domain superfamily (SSF46785).

HAstV Replication Cycle

Replication of genomes of all well-characterized positive-strand RNA viruses — including plant, animal, and insect viruses — occurs in large

replication complexes associated with intracellular membranes. For HAstV, it has been suggested that RNA replication is associated to membranes derived from the endoplasmic reticulum, at the perinuclear region, and ultrastructural analysis has revealed the presence of viral aggregates close to the nuclear periphery surrounded by a high number of double-membrane vacuoles.[47] Although the exact composition of these replication complexes is not known, it is likely that all nonstructural proteins are required and that the temporary regulation of protein processing of the replicase complex may play a role in the regulation of the minus- and plus-strand RNA synthesis processes. Of interest, variability within the protein encoded at the C-terminal region of ORF1a has been shown to affect the RNA replication pattern, and a certain functional effect of phosphorylation of the protein has been suggested.[51] Although some nonstructural proteins have been detected by immunofluorescence within the nucleus of infected cells,[48] the involvement of the nuclear phase in the replication cycle of HAstV is not fully understood.

On the other hand, the coordination of genome packaging into maturing particles has not yet been defined, but the structural determinants for particle assembly seem to be carried out by the structural proteins alone, since it is possible to make VLPs even without the first 70 amino acids of the capsid polyprotein. The fact that structural proteins colocalize with nonstructural proteins suggests the possibility that the membranes might form a scaffold for packaging of the viral RNA into assembling capsids.

Finally, transient expression experiments showed a direct link between HAstV ORF1a encoded proteins and apoptosis induction.[28] It seems that astrovirus replication is tightly linked to an apoptotic host cell response, since cellular caspases seem to be responsible for the first step of capsid maturation, before the trypsin cleavages that occur extracellularly.[23] While a premature apoptotic response usually impairs viral infectivity, when enough virus progeny has been generated, apoptosis may facilitate the release of virus to bystander cells, evading the inflammatory response. Kinetic analysis of viral replication, indicated by the synthesis of structural proteins from the subgenomic RNA, and onset of apoptosis showed that host cell

apoptosis would be triggered subsequent to viral replication,[28] suggesting that the apoptotic process is a late event in the replication cycle of HAstV.

Future Challenges

Although HAstV infectious particles were discovered almost 30 years ago, limited information is still available on the molecular characteristics of the virus, and its genome organization has not been completely clarified. However, the availability of molecular techniques to study epidemiological features of viral infection has resulted in an extensive database of sequence information, with more than 400 entries in the Genbank, and we are beginning to understand its burden in gastroenteritis worldwide, the extent of its variability, and the main characteristics of immunologic responses to infection. Aside from the use of some cell-adapted HAstV strains that replicate in CaCo-2 cells to high titers, the construction of an infectious full-length clone with serotype 1 has been a critical step in development of genetic systems for study of HAstV.[16] Not only may it be useful for identifying the cleavage sites of both nonstructural polyproteins and the capsid precursor, as well as its biological functions, but it also offers promise as a paradigm for understanding the interactions between viral and host cellular processes. Studies on HAstV capsid structure will provide new insights into the understanding of HAstV genetic and antigenic diversity, as well as the characterization of the immune response, two key steps before the generation of potential vaccines.

Acknowledgments

S. Guix was recipient of an FI fellowship from the Generalitat de Catalunya. We acknowledge the technical expertise of the Serveis Científic-Tècnics of the University of Barcelona. This work was supported in part by grants SP22-CT-2004-502571 from the European Union, 2001/SGR/00098 from the Generalitat de Catalunya, and

the Centre de Referència de Biotecnologia de Catalunya (CeRBa), Generalitat de Catalunya.

References

1. World Health Report. (2004) WHO, Geneva.
2. Taterka JA, Cuff CF, Rubin DH. (1992) Viral gastrointestinal infections. *Gastroenterol Clin North Am* **21**: 303–330.
3. Glass RI, Noel J, Mitchell D, *et al.* (1996) The changing epidemiology of astrovirus-associated gastroenteritis: a review. *Arch Virol (Suppl)* **12**: 287–300.
4. Matsui SM, Greenberg HB. (2001) Astroviruses. In B.N. Fields, D.M. Knipe, P.M. Howley, *et al.* (eds), *Fields virology*, pp.875–893. Lippincott Williams and Wilkins, Philadelphia, Pennsylvania.
5. Walter JE, Mitchell DK. (2003) Astrovirus infection in children. *Curr Opin Infect Dis* **16**: 247–253.
6. Appleton H, Higgins PG. (1975) Viruses and gastroenteritis in infants. *Lancet* **1**: 1297.
7. Madeley CR, Cosgrove BP. (1975) 28 nm particles in faeces in infantile gastroenteritis. *Lancet* **2**: 451–452.
8. Méndez E, Fernández-Luna T, López S, *et al.* (2002) Proteolytic processing of a serotype 8 human astrovirus ORF2 polyprotein. *J Virol* **76**: 7996–8002.
9. Méndez-Toss M, Romero-Guido P, Munguia ME, *et al.* (2000) Molecular analysis of a serotype 8 human astrovirus genome. *J Gen Virol* **81**: 2891–2897.
10. Taylor MB, Walter JE, Berke T, *et al.* (2001) Characterisation of a South African human astrovirus as type 8 by antigenic and genetic analyses. *J Med Virol* **64**: 256–261.
11. Jiang B, Monroe SS, Koonin EV, *et al.* (1993) RNA sequence of astrovirus: distinctive genomic organization and a putative retrovirus-like ribosomal frameshifting signal that directs the viral replicase synthesis. *Proc Natl Acad Sci USA* **90**: 10539–10543.
12. Lewis TL, Greenberg HB, Herrmann JE, *et al.* (1994) Analysis of astrovirus serotype 1 RNA, identification of the viral RNA-dependent RNA polymerase motif, and expression of a viral structural protein. *J Virol* **68**: 77–83.

13. Oh D, Schreier E. (2001) Molecular characterization of human astroviruses in Germany. *Arch Virol* **146**: 443–455.
14. Willcocks MM, Brown TD, Madeley CR, Carter MJ. (1994) The complete sequence of a human astrovirus. *J Gen Virol* **75**: 1785–1788.
15. Monroe SS, Stine SE, Gorelkin L, *et al.* (1991) Temporal synthesis of proteins and RNAs during human astrovirus infection of cultured cells. *J Virol* **65**: 641–648.
16. Geigenmüller U, Ginzton N, Matsui SM. (1997) Construction of a genome-length cDNA clone for human astrovirus serotype 1 and synthesis of infectious RNA transcripts. *J Virol* **71**: 1713–1717.
17. Al-Mutairy B, Walter JE, Pothen A, Mitchell DK. (2005) Genome prediction of putative genome-linked viral protein (VPg) of astroviruses. *Virus Genes* **31**: 21–30.
18. Sosnovtsev S, Green KY. (1995) RNA transcripts derived from cloned full-length copy of the feline calicivirus genome do not require VPg for infectivity. *Virology* **210**: 383–390.
19. Matsui SM, Kiang D, Ginzton N, *et al.* (2001) Molecular biology of astroviruses: selected highlights. *Novartis Foundation Symposium* **238**: 219–233.
20. Jonassen CM, Jonassen TO, Sveen TM, Grinde B. (2003) Complete genomic sequences of astroviruses from sheep and turkey: comparison with related viruses. *Virus Res* **91**: 195–201.
21. Walter JE, Briggs J, Guerrero ML, *et al.* (2001) Molecular characterization of a novel recombinant strain of human astrovirus associated with gastroenteritis in children. *Arch Virol* **146**: 2357–2367.
22. Bass DM, Qiu S. (2000) Proteolytic processing of the astrovirus capsid. *J Virol* **74**: 1810–1814.
23. Méndez E, Salas-Ocampo E, Arias CF. (2004) Caspases mediate processing of the capsid precursor and cell relase of human astroviruses. *J Virol* **78**: 8601–8608.
24. Sánchez-Fauquier A, Carrascosa AL, Carrascosa JL, *et al.* (1994) Characterization of a human astrovirus serotype 2 structural protein (VP26) that contains an epitope involved in virus neutralization. *Virology* **201**: 312–320.
25. Dalton RM, Pastrana EP, Sánchez-Fauquier A. (2003) Vaccinia virus recombinant expressing an 87-kilodalton polyprotein that is sufficient to form astrovirus-like particles. *J Virol* **77**: 9094–9098.

26. Geigenmüller U, Ginzton N, Matsui SM. (2002) Studies on intracellular processing of the capsid protein of human astrovirus serotype 1 in infected cells. *J Gen Virol* **83**: 1691–1695.
27. Caballero S, Guix S, Ribes E, *et al.* (2004) Structural requirements of astrovirus virus-like particles assembled in insect cells. *J Virol* **78**: 13285–13292.
28. Guix S, Bosch A, Ribes E, *et al.* (2004) Apoptosis in astrovirus-infected CaCo-2 cells. *Virology* **319**: 249–261.
29. Al-Molawi N, Beardmore VA, Carter MJ, *et al.* (2003) Caspase-mediated cleavage of the feline calicivirus capsid protein. *J Gen Virol* **84**: 1237–1244.
30. Best SM, Shelton JF, Pomepuy JM, *et al.* (2003) Caspase cleavage of the non-structural protein NS1 mediates replication of Aleutian mink disease virus. *J Virol* 77: 5305–5312.
31. Eleouet J-F, Slee EA, Saurini F, *et al.* (2000) The viral nucleocapsid protein of transmissible gastroenteritis coronavirus (TGEV) is cleaved by caspase-6 and -7 during TGEV-induced apoptosis. *J Virol* **74**: 3975–3983.
32. Goh PY, Tan YJ, Lim SP, *et al.* (2001) The hepatitis C virus core protein interacts with NS5A and activates its caspase-mediated proteolytic cleavage. *Virology* **290**: 224–236.
33. Grand R, Schmeiser K, Gordon E, *et al.* (2002) Caspase-mediated cleavage of adenovirus early region 1A proteins. *Virology* **301**: 255–271.
34. Zhirnov OP, Konakova TE, Garten W, Klenk HD. (1999) Caspase-dependent N-terminal cleavage of influenza virus nucleocapsid protein in infected cells. *J Virol* **73**: 10158–10163.
35. Krishna NK. (2005) Identification of structural domains involved in astrovirus capsid biology. *Viral Immunol* **18**: 17–26.
36. Wang Q, Kakizawa J, Wenng L, *et al.* (2001) Genetic analysis of the capsid region of astroviruses. *J Med Virol* **64**: 245–255.
37. Bass DM, Upadhyayula U. (1997) Characterization of human serotype 1 astrovirus-neutralizing epitopes. *J Virol* **71**: 8666–8671.
38. Herrmann JE, Hudson RW, Perron-Henry DM, *et al.* (1988) Antigenic characterization of cell-cultivated astrovirus serotypes and development of astrovirus-specific monoclonal antibodies. *J Infect Dis* **158**: 182–185.

39. Baer ML, Houser F, Loesch-Fries LS, Gehrke L. (1994) Specfic RNA binding by amino-terminal peptides of alfalfa mosaic virus coat protein. *EMBO J* **13**: 727–735.
40. Geigenmüller U, Chew T, Ginzton N, Matsui SM. (2002) Processing of nonstructural protein 1a of human astrovirus. *J Virol* **76**: 2003–2008.
41. Geigenmüller-Gnirke U, Nitschko H, Schlesinger S. (1993) Deletion analysis of the capsid protein of Sindbis virus: identification of the RNA binding region. *J Virol* **67**: 1620–1626.
42. Risco C, Carrascosa JL, Pedregosa AM, *et al.* (1995) Ultrastructure of human astrovirus serotype 2. *J Gen Virol* **76**: 2075–2080.
43. Kelley LA, MacCallum RM, Sternberg MJE. (2000) Enhanced Genome Annotation using Structural Profiles in the Program 3D-PSSM. *J Mol Biol* **299**: 499–520.
44. Harrison SC. (2001) Principles of virus structure. In B.N. Fields, D.M. Knipe, P.M. Howley, *et al* (eds), *Fields virology*, pp. 53–85. Lippincott-Raven Publishers, Philadelphia, Pennsylvania.
45. White LJ, Hardy ME, Estes MK. (1997) Biochemical characterization of a smaller form of recombinant Norwalk virus capsids assembled in insect cells. *J Virol* **71**: 8066–8072.
46. Kiang D, Matsui SM. (2002) Proteolytic processing of a human astrovirus nonstructural protein. *J Gen Virol* **83**: 25–34.
47. Guix S, Caballero S, Bosch A, Pintó RM. (2004) C-terminal nsP1a protein of Human Astrovirus colocalizes with the endoplasmic reticulum and viral RNA. *J Virol* **78**: 13627–13636.
48. Willcocks MM, Boxall AS, Carter MJ. (1999) Processing and intracellular location of human astrovirus non-structural proteins. *J Gen Virol* **80**: 2607–2611.
49. Matsui SM, Kim JP, Greenberg HB, *et al.* (1993) Cloning and characterization of human astrovirus immunoreactive epitopes. *J Virol* **67**: 1712–1715.
50. Willcocks MM, Ashton N, Kurtz JB, *et al.* (1994) Cell culture adaptation of astrovirus involves a deletion. *J Virol* **68**: 6057–6058.
51. Guix S, Caballero S, Bosch A, Pintó RM. (2005) Human astrovirus C-terminal nsP1a protein is involved in RNA replication. *Virology* **333**: 124–131.
52. Dunham DM, Jiang X, Berke T, *et al.* (1998) Genomic mapping of a calicivirus VPg. *Arch Virol* **143**: 2421–2430.

53. Gibson CA, Chen J, Monroe SA, Denison MR. (1998) Expression and processing of nonstructural proteins of the human astroviruses. *Adv Exp Med Biol* **440**: 387–391.
54. Méndez E, Salas-Ocampo MP, Munguía ME, Arias CF. (2003) Protein products of the open reading frames encoding nonstructural proteins of human astrovirus serotype 8. *J Virol* **77**: 11378–11384.
55. Grove A, Wilkinson SP. (2005) Differential DNA binding and protection by dimeric and dodecameric forms of the ferritin homolog Dps from Deinococcus radiodurans. *J Mol Biol* **347**: 495–508.
56. Muller AJ, Baker JF, DuHadaway JB, *et al.* (2003) Targeted disruption of the murine Bin1/Amphiphysin II gene does not disable endocytosis but results in embryonic cardiomyopathy with aberrant myofibril formation. *Mol Cell Biol* **23**: 4295–4306.
57. Barrette-Ng IH, Ng KK, Mark BL, *et al.* (2002) Structure of arterivirus nsp4. The smallest chymotrypsin-like proteinase with an alpha/beta C-terminal extension and alternate conformations of the oxyanion hole. *J Biol Chem* **18**: 39960–39966.
58. Jeffery CJ. (2004) Molecular mechanisms for multitasking: recent crystal structures of moonlighting proteins. *Curr Opin Struct Biol* **14**: 663–668.
59. Jonassen CM, Jonassen TO, Saif YM, *et al.* (2001) Comparison of capsid sequences from human and animal astroviruses. *J Gen Virol* **82**: 1061–1067.

This page intentionally left blank

Chapter 23

DNA Vaccines against Viruses

Britta Wahren[†] *and Margaret A. Liu*[*,†,‡]

Rationale for Gene-based Vaccines

The administration of genes encoding antigens, either via viral or bacterial vectors, or as plasmid DNA, may be the next generation of vaccines and therapeutics to arise from recombinant DNA technology. Although vaccines have been amazingly effective for a variety of diseases, other diseases have been less amenable to prevention or control through traditional vaccine approaches. The rationale for using genes rather than proteins or forms of the pathogens themselves arose from a variety of immunologic, manufacturing and safety challenges unique to the pathogens (such as HIV and malaria) or the disease (such as cancer, autoimmune diseases, or allergy). The particular characteristics of DNA plasmids in terms of the immune responses that are generated against the antigens encoded by the DNA, and the features of DNA as a product, relative to other biologicals, have led

* Corresponding author.
† Department of Microbiology and Tumor and Cell Biology, Karolinska Institute, Swedish Institute for Infectious Disease Control, Stockholm, Sweden.
‡ ProTherImmune, Lafayette, California, USA.

to their widespread development and testing for prevention and treatment of a variety of diseases.

DNA vaccines were a surprising development from efforts to formulate bacterial plasmid DNA for uptake by cells *in vivo*. Because viruses have evolved complex structures and mechanisms in order to deliver their genetic material into cells, it was long thought that so-called "naked," unformulated DNA would not be useful as a means to directly deliver genes into cells *in vivo*. Thus the observation by Felgner and colleagues that bacterial plasmid DNA suspended simply in saline could result in the uptake of the DNA (albeit relatively inefficiently) by muscle cells *in vivo* — with the subsequent transcription and translation of the encoded genes — came as a surprise.[1] Just as unexpected was the ability of the DNA to result in the generation of a cytotoxic T lymphocyte response that was effective at protecting against a viral challenge,[2] because myocytes are not professional antigen-presenting cells (APC), and presentation of antigen by a non-professional APC usually results in tolerance or non-response rather than activated cytolytic T cells. The effectiveness of the DNA for the induction of immune responses was striking because the DNA was able to induce both antibodies and T cells against the encoded antigen. And as noted above, the cellular immune response was potent enough to protect against death (but not disease) from a viral challenge. Moreover, the challenge virus was of a strain different from the strain from which the antigen gene had been cloned (i.e. it was heterosubtypic). This type of cross-strain immunity is one of the great challenges of vaccine development for diseases such as influenza and HIV where the virus mutates frequently and easily enough such that antibody responses against one strain are ineffective against another strain. In the case of influenza this means that the vaccine must contain several different strains of virus, and the vaccine composition must be made new each year to provide the strains that are judged to be the currently circulating strains. In the case of HIV, the tremendous variability of the virus is one of the major reasons that no vaccine has been made to date.

A number of other rationales exist for delivery of proteins in the form of genes rather than as proteins. For example, cytokine administration

has been one of the great successes of recombinant DNA technology, such as the use of recombinant cytokines for the treatment of cancers and chronic viral infections. Yet it is clear that administering the gene rather than the protein could have numerous advantages: proteins synthesized *in situ* could avoid the toxicities associated with systemic administration; administration of genes could mean that the protein expression could persist for longer periods of time (compared to the short half-lives of the recombinant proteins); and a protein synthesized *in situ* would have appropriately mammalian post-translational modifications, thus circumventing one of the major challenges that can occur when producing recombinant proteins in non-mammalian host cells.

DNA vaccines have further potential advantages as a product, besides the immunologic and clinical issues noted above. The manufacturing process is relatively generic compared to either small molecule drugs or other biologicals such as either recombinant proteins or live viral vaccines. DNA vaccines are bacterial plasmids, with the differences generally being simply the gene insert and a promoter. So the production process of growth in bacterial hosts and the subsequent purification of the plasmids is reasonably similar for different vaccines or therapies, and the plasmids are easier to purify than recombinant proteins. Moreover, DNA vaccines are more stable than live viruses. Current embodiments of DNA vaccine include formulations and delivery systems that may make them more complicated as products,[3,4] but the generic nature of the entity remains a compelling attribute.

Pre-Clinical Proof of Concept

The demonstration that DNA vaccines could work pre-clinically *in vivo* was done using a simple plasmid of DNA coding for a protein from the influenza virus. That study showed that the DNA could result in the generation of both cytotoxic T lymphocytes and antibodies, and more importantly, could protect from an otherwise lethal challenge with a strain of influenza different from the strain from which the gene had been cloned.[2] The ability to protect the animals in a

cross-strain manner raised considerable interest because it offered a potential means to design vaccines for diseases such as HIV, influenza, and hepatitis C, whose multiple and/or evolving strains pose such a challenge for vaccinologists. While the current influenza vaccine cannot protect against a virus of the same subtype that has drifted even a minor amount, the DNA vaccine protected against a challenge from a completely different subtype. This raised hopes that DNA vaccines would provide a new means for making vaccines that would be broadly effective against a variety of strains, thereby helping to prevent the types of global epidemics such as the 1919 influenza pandemic that killed about 20 million people, or be a foundation for a rapidly produced component against bird influenza, H5N1. The hope was also that the DNA vaccines would be a means to make vaccines not only against other viral diseases such as HIV, but also against other viruses whose sequellae included cancer (such as papilloma virus[5] and hepatitis C) and infectious diseases such as malaria that would benefit from the cytotoxic T cell response.

During the last few years, progress has been made in the area of protection against both viral and bacterial diseases. Recent expression analysis has shown that DNA antigens may act as endogenous immune activators in specific species.[6,7] Also, the envelope antigens of two acute RNA viruses (rhabdo and filo) have been shown to confer protective capacity after vaccination with their DNA homologs. The single DNA representing the G protein of rhabdovirus was shown early to protect mice. Envelope genes representing glycoproteins of the Ebola virus strains from central Africa (Zaire and Sudan), together with the conserved nucleoprotein gene induced antibody in all the healthy subjects, while one third of the immunized persons obtained cytolytic T-cell reactivity.[8] Antibody is assumed to be the most important protective entity for acute RNA viruses, while cell-mediated immunity is necessary for infections establishing persistent infections. Plasmid DNA encoding HIV proteins incorporated in a prime-boost schedule predicted the potent and broad immunogenicity subsequently found in humans.[3]

Additional studies in a variety of pre-clinical models of various diseases rapidly demonstrated the broad applicability of the DNA vaccine technology. Perhaps just as importantly, the ease with which the various laboratories used off-the-shelf plasmids and bacterial lines for production of the DNA vaccines demonstrated the robustness of the technology. The models in which DNA vaccines were utilized for inducing immune responses or pre-clinical efficacy included a variety of infectious diseases caused by viruses, bacteria and parasites. Other types of disease such as cancer, autoimmune diseases and allergy[9,10] were also shown to be amenable to DNA vaccine prophylaxis or therapy in pre-clinical models. The applications for allergies (asthma)[11] and autoimmune diseases (such as diabetes)[12] are based upon the ability of the DNA to not simply generate cellular immunity, but to alter the type of T cell helper response for the particular protein antigen. This shift in the Th1 and Th2 responses and other data demonstrate that the role of the DNA is not simply to function as a carrier of the gene encoding the antigen, but that the DNA itself also has immune effects and/or that the method of generation of the immune response is different from the means by which the abnormal autoimmune allergic response against a given antigen occur.

A number of applications have moved the field forward regarding the influence on allergic diseases, autoimmune diseases and cancer. One of the most interesting properties of a DNA vaccine, is its ability to not only induce Th1 responses, but to also divert the immune response away from antibody production. Thus, allergic responses by IgE have been redirected by small doses of DNA encoding endogenous enzymes, such as the beta-galactosidase. The anti-allergic properties of the DNA plasmid were attributed both to an induction of Th1 response, as well as a non-induction of IgE antibodies, which induce allergy in this preclinical mouse model.[13] The balance between allergic disease and autoimmunity is fine, however, and in some diseases an autoimmune reactivity is instead desirable. This approach involves the induction by DNA of immune responses against slightly heterologous substances.

Studies have addressed the means by which the DNA is able to generate cytolytic T cell responses despite the concern that protein

expression in the non-antigen presenting cells (i.e. the myocyte) would result in immune tolerance or non-responsiveness. Several groups have demonstrated that antigen produced in the muscle can be transferred into professional antigen-presenting cells by a method known as cross-priming such that cytolytic T cells are generated.[14–16] Additional studies revealed that the vector backbone of the plasmid, because it is bacterial DNA with a different methylation pattern than mammalian DNA, stimulates the innate immune system via a Toll-Like Receptor, thus potentially affecting the immunogenicity of the DNA vaccine and increasing its potency.[17]

Nevertheless, despite these novel mechanisms contributing to the immunogenicity of the DNA vaccines, it became apparent that, while robust in a variety of animal models, DNA vaccines require larger doses or different adjuvants in non-human primates and humans at least partly since their innate sensors are different.[4,5,9,10,18,19] These observations have resulted in a variety of modifications to make the DNA vaccines more potent.

Licensed Vèterinary DNA Vaccines

In 2005, the first veterinary DNA-based vaccines were licensed, for salmon against infectious hematopioetic necrosis virus (Apex-IHN, Novartis) and for horses against West Nile virus infection (Fort Dodge Laboratories).

A first licensed DNA cancer vaccine for melanoma in dogs encodes molecules from the tyrosine kinase family. Immunization with foreign TYR DNA improved survival and even cured dogs with progressing melanoma. In this case, the heterogenous tyrosinase was able to break tolerance against the endogenous enzyme.[20] Antibodies appeared to be the major effector molecules occurring in parallel with tumor rejection. A similar approach has been initiated for human melanoma. Patients with melanoma in stages 3–4 have received the murine tyrosinase DNA followed by human TYR DNA. Half of the patients have developed cytolytic $CD8^+$ cells directed against a TYR peptide. The median survival time has not yet been reached, indicating a prolonged survival.[21]

Clinical Trials

DNA vaccines encoding antigens from pathogens and tumors have been tested in early phase human trials. The initial studies were focused on demonstrating the safety of this new class of vaccine. Although older, widely utilized vaccines in some cases contained bacterial DNA (for example, the earlier version of the pertussis vaccine, which was made of killed pertussis), the use of recombinant DNA — although highly purified with a known sequence of the DNA — required extensive pre-clinical safety testing as well as the usual phase I clinical safety studies. In various studies for diseases ranging from HIV and influenza to the parasite malaria, the clinical trials generally demonstrated weak but real immune responses, both antibodies and cellular immune responses.[22,23] The low level of the responses was disappointing given the ease with which potent immunity had been demonstrated in such a large number of animal models. Yet certain interesting observations have emerged. For example, certain HIV patients who had had long exposures to high levels of viral antigens (due to their high viral loads), yet who had not made cytolytic T cell responses against a particular viral antigen, mounted cytolytic T cell responses against epitopes of that protein following the DNA immunization.[22–24] Thus, by some mechanism, the DNA resulted in the generation of immune responses not evoked by the virus. This provided evidence that different methods of immunization (if infection with the virus to induce immune responses may be called immunization) can have different outcomes in terms of immune responses.

New Generation DNA Vaccines

Efforts have focused on ways to increase the potency of DNA vaccines. One early and continuing approach has been to redesign the plasmids to enable them to produce more protein antigen.[25,26] As data from early clinical trials demonstrated the safety of administering DNA vaccines to humans, increased doses of DNA have been used in clinical trials, increasing to as much as 1–8 milligram doses per vaccine.[27,28] Other efforts have been directed at formulating the DNA in order to

increase its cellular uptake, or to enable it to withstand extracellular degradation for longer. Encapsulating the DNA into[29] or adhering the DNA onto[30] microparticles likewise appears to increase the potency of DNA vaccines by either protecting the DNA from degradation or increasing its uptake by antigen processing cells, or both. Variants of DNA have been incorporated into the sequences, such as the nt sequence for anticipated drug mutations.[31]

Genes encoding immune-stimulating molecules such as cytokines have been added to the plasmids, in essence adding adjuvants to the genes encoded by the DNA.[32] As noted above, the bacterial origin of the plasmids means that the sequence and hence the methylation pattern of the DNA enables it to activate Toll-like receptors, potentially engaging the innate immune system in the process of stimulating cognate immunity.[33]

Different delivery devices or routes have also been employed in efforts to increase the potency of the DNA vaccines, to stimulate specific types of immunity (such as mucosal immune responses), or to facilitate delivery for patient compliance or for resource-poor settings. The initial demonstration of a DNA vaccine that could stimulate antibody responses employed a "gene gun" to shoot DNA-coated gold beads into the skin.[34] Such "particle-mediated" injection has been used in clinical trials to generate antibodies against Hepatitis B surface antigen[35] using DNA encoding the antigen. Although the DNA immunizations resulted in lower titers and utilized more immunizations that with the licensed protein vaccine, they still showed the ability of DNA immunized via the gold beads to clinically produce the desired immune response. In a further study, patients who had not responded well to the licensed (recombinant protein) hepatitis vaccine made antibodies following immunization with the DNA vaccine[36] given by gene gun.

DNA transfer by apoptotic cells appears to induce a receptor-independent immune response. The DNA, present in the apoptotic bodies from transfection by plasmids or by the infectious agent, is taken up by antigen presenting cells and appears to induce good immune responses.[37]

Devices other than traditional syringes are being tested to directly propel the DNA vaccine into the skin (for example, the Biojector)[3,38]

or mucosa (for example the Syriject).[39] Another approach has been to couple the usual injection with subsequent *in vivo* electroporation (resulting in an increase in the amount of transfection of the DNA into cells).[4,40,41]

Trials in non-human primates and humans are evaluating the novel and promising concept of using the DNA vaccine as a prime to be followed by immunization with viral vectors encoding the same antigen or a recombinant protein version of the antigen.[42–46] This approach has been termed the "prime-boost" concept. The immunologic mechanism for the potency of these sequences is not known, but a variety of vectors used as the boost appear to be useful. Of note, the immune responses are most potent when the DNA vaccine is the priming agent.

Conclusions

DNA vaccines hold promise as a means to generate the types of immunity needed for a variety of diseases, including either via the induction of cross-strain cellular immune responses or the modulation of the type of immunity such as the type of T cell help. In addition, even though modifications of DNA vaccines by the addition of formulations or delivery devices have made DNA vaccines more complex than their originally envisioned simple plasmid in saline, the vaccine still holds appeal because the relatively generic nature of the plasmid. But the most compelling rationale for DNA vaccines is to offer the potential for making vaccines and therapeutics directed against diseases that to date have not been effectively addressed by earlier technologies, including chronic infections, diseases due to immunological aberrations, and tumors.

References

1. Wolff JA, Malone RW, Williams P, *et al.* (1990) *Science* **247**: 1465–1468.
2. Ulmer JB, Donnelly JJ, Parker SE, *et al.* (1993) *Science* **259**: 1745–1749.
3. Bråve A, Boberg A, Gudmundsdotter L, *et al.* (2007) *Molecular Therapy* **15**: 1724–1733.
4. Luckay A, Sidhu MK, Kjeken R, *et al.* (2007) *J Virol.* **81**: 5257–5269.

5. Donnelly JJ, Martinez D, Jansen KU, *et al.* (1996) *J Infect Dis* **173**: 314–320.
6. Purcell MK, Nichols KM, Winton JR, *et al.* (2006) *Mol Immunol* **43**: 2089–2106.
7. Yasuike M, Kondo H, Hirono I, *et al.* (2007) *Fish Shellfish Immunol* **23**: 531–541.
8. Martin JE, Sullivan MJ, Enama ME, *et al.* (2006) *Clin Vaccine Immunol* **13**: 1267–1277.
9. Srivastava IK, Liu MA. (2003) *Ann Int Med* **138**: 550–559.
10. Liu MA. (2003) *J Intern Med* **253**: 402–410.
11. Jarman ER, Lamb JR. (2004) *Immunology* **112**: 631–642.
12. Prud'homme GJ. (2003) *Expert Rev Vaccines* **2**: 533–540.
13. Scheiblhofer S, Gabler M, Leitner WW, *et al.* (2006) *Allergy* **61**: 828–835.
14. Fu TM, Ulmer JB, Caulfield MJ, *et al.* (1997) *Molecular Medicine* **3**: 362–371.
15. Torres CA, Iwasaki A, Barber BH, *et al.* (1997) *J Immunol* **158**: 4529–4532.
16. Doe B, Selby M, Barnett S, *et al.* (1996) *Proc Natl Acad Sci USA* **93**: 8578–8583.
17. Sato Y, Roman M, Tighe H, *et al.* (1996) *Science* **273**: 299–302.
18. Klinman DM, Xie H, Ivins BE. (2006) *Ann N Y Acad Sci* **1082**: 137–150
19. Gu M, Hine PM, James Jackson W, *et al* (2007) *Vaccine* **25**: 526–534.
20. Bergman PJ, Camps-Palau MA, McKnight JA, *et al.* (2006) *Vaccine* **24**: 4582–4585.
21. Wolchok JD, Yuan J, Houghton AN, *et al.* (2007) *Mol Ther Aug* 28, Epub.
22. Calarota S, Bratt G, Nordlund S, *et al.* (1998) *Lancet* **351**: 1320–1325.
23. MacGregor RR, Boyer JD, Ugen KE, *et al.* (1998) *J Inf Dis* **178**: 92–101.
24. Calarota SA, Kjerrstrom A, Islam KB, *et al.* (2001) *Hum Gene Ther* **12**: 1623–1637.
25. zur Megede J, Chen MC, Doe B, *et al.* (2000) *J Virol* **74**: 2628–2635.
26. André S, Seed B, Eberle J, *et al.* (1998) *J Virol* **72**: 1497–1503.
27. Graham BS, Koup RA, Roederer M, *et al.* (2006) *J Infect Dis* **194**: 1650–1660.
28. Le TP, Coonan KM, Hedstrom RC, *et al.* (2000) *Vaccine* **18**: 1893–1901.
29. Howard KA, Li XW, Somavarapu S, *et al.* (2004) *Biochim Biophys Acta* **1674**: 149–157.
30. O'Hagan DT, Singh M, Ulmer JB. (2004) *Immunol Rev* **199**: 191–200.

31. Boberg A, Gaunitz S, Bråve A, *et al.* (2008) Vaccine in press.
32. Chong SY, Egan MA, Kutzler MA, *et al.* (2007) Vaccine **25**: 4967–4982.
33. Klinman DM, Yamshchikov G, Ishigatsubo Y. (1997) *J Immunol* **158**: 3635–3639.
34. Tang DC, DeVit M, Johnston SA. (1992) *Nature* **356**: 152–154.
35. Roy MJ, Wu MS, Barr LJ, *et al.* (2000) *Vaccine* **19**: 764–778.
36. Rottinghaus ST, Poland GA, Jacobson RM, *et al.* (2003) *Vaccine* **21**: 4604–4608.
37. Spetz AL, Sorensen AS, Walther-Yallow L, *et al.* (2002) *J Immunol* **169**: 5771–5779.
38. Trimble C, Lin CT, Hung CF, *et al.* (2003) *Vaccine* **21**: 4036–4042.
39. Lundholm P, Leandersson AC, Christensson B, *et al.* (2002) *Virus Res* **82**: 141–145.
40. Otten G, Schaefer M, Doe B, *et al.* (2004) *Vaccine* **22(19)**: 2489–2493.
41. Roos AK, Moreno S, Leder C, *et al.* (2006) *Mol Ther* **13**: 320–327.
42. Schneider J, Gilbert SC, Blanchard TJ, *et al.* (1998) *Nat Med* **4**: 397–402.
43. Moorthy VS, Imoukhuede EB, Keating S, *et al.* (2004) *J Infect Dis* **189**: 2213–2219.
44. Epstein JE, Charoenvit Y, Kester KE, *et al.* (2004) *Vaccine* **22**: 1592–1603.
45. Sandström E, Wahren B, Hejdeman B, *et al* (2007). *AIDS conference* Seattle, Abstr.
46. Amara RR, Villinger F, Altman JD, *et al.* (2001)*Science* **292**: 69–74.

This page intentionally left blank

Chapter 24

Life Cycles of Polyomaviridae — DNA Tumor Virus

Masaaki Kawano,[†] *Hiroshi Handa*[‡] *and R. Holland Cheng*[*,†]

Simian virus 40 (SV40) is a non-enveloped DNA virus. It is known to have been accidentally introduced to a large segment of the human population during the mid-20th century through tainted polio vaccines produced in the kidney cells of African green monkeys and rhesusu monkeys. Studies have indicated that when the virus invades certain mammalian cells, its reproduction is often hindered, but the virus may produce T-antigen and incite oncogenic transformation in the host. SV40 has been shown to incite tumor growth in rodents and may be linked to the development of mesothelioma and certain other kinds of cancer in humans. Therefore, SV40 has been well analyzed as a tumor model. Recently, the host cell receptors for SV40 have been identified and it has been shown to pass from surface caveolae to the endoplasmic reticulum in apparently novel infectious entry pathway.

*E-mail: rhch@ucdavis.edu

[†]Molecular and Cellular Biology, University of California, Davis CA 95616, USA.

[‡]Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8503, Japan.

Abbreviations: Simian virus 40; SV40; Virus-like particles; VLPs; polyomavirus; JC; BK; minichromosome; Nuclear localization signals; NLSs; Sp1; hsc70; DnaK; J-domain; large T-antigen; hsp70; the major histocompatibility complex; MHC; human lymphocyte antigen; HLA; integrin; neuraminidase; gangliosides; GD1a; GT1b; GM1; GD1b; endocytosis; caveolae; endoplasmic reticulum; ER; EGTA; dithiothreitol; DTT; (PDI)-like protein; ERp29; gene transfer; Propidium iodide; PI; human dendritic cells; DC; vaccination.

Virus-like particles (VLPs) of polyomavirus has been shown to enter into cells in the same way as polyomavirus virion, suggesting the possibility that VLPs can be used as the carrier of drug delivery system through the use of the polyomavirus infectious pathway.

Virion Structure

The capsid of SV40, like those of other small double-stranded DNA viruses in the polyomavirus family (polyomavirus, JC, and BK), is composed of major capsid proteins, VP1, arranged on a T = 7d icosahedral lattice (about 500 Å in diameter).[1–3] They can induce tumors in laboratory animals and transform cells in culture. The virus particles deliver a minichromosome from the nucleus of one cell to the nucleus of another. They have been studied crystallographically, and they are icosahedrally symmetrical particles, containing 360 copies of a major structural protein, VP1, arranged as 72 VP1-pentameric building blocks on the viral surface (Figs. 1A and 1B). Twelve VP1-pentamers lie on the twelve 5-fold rotation axes of the icosahedron, each surrounded by five other VP1-pentamers (Fig. 2A, light blue). The remaining 60 VP1-pentamers do not lie on the symmetry axes, and they are surrounded by six other VP1-pentamers (Fig. 2A, white). Long C-terminal arms of the VP1 molecule tie the VP1-pentamers together; they extend from one VP1-pentamer and fit into binding sites on adjacent VP1-pentamers. The 5- and 6-coordinated VP1-pentamers lead to just three kinds of interpentamer contacts: an approximate 3-fold axis relating to subunits α, α' and α'' (Fig. 2B); an approximate 2-fold axis relating to β and β' (Fig. 2C); and a strict icosahedral 2-fold axis relating to γ(Fig. 2D). The study of the *in vitro* particle formation of the polyomavirus major structural VP1 protein purified after expression of the recombinant gene in *E. coli* showed the association of the capsid-like assemblies and polymorphic aggregates at high ionic strength.[4] SV40 VP1-pentamers prepared from VLPs produced in Sf9 cells selectively formed capsid-like assemblies in the pressence of DNA in physiologically conditions *in vitro*. These nonequivalent assemblies related VP1 subunits of the penta- and hexavalent capsomers, therefore, must spontaneously switch their bonding specificity during assembly. VP1 can switch inter VP1 bonding specificity in

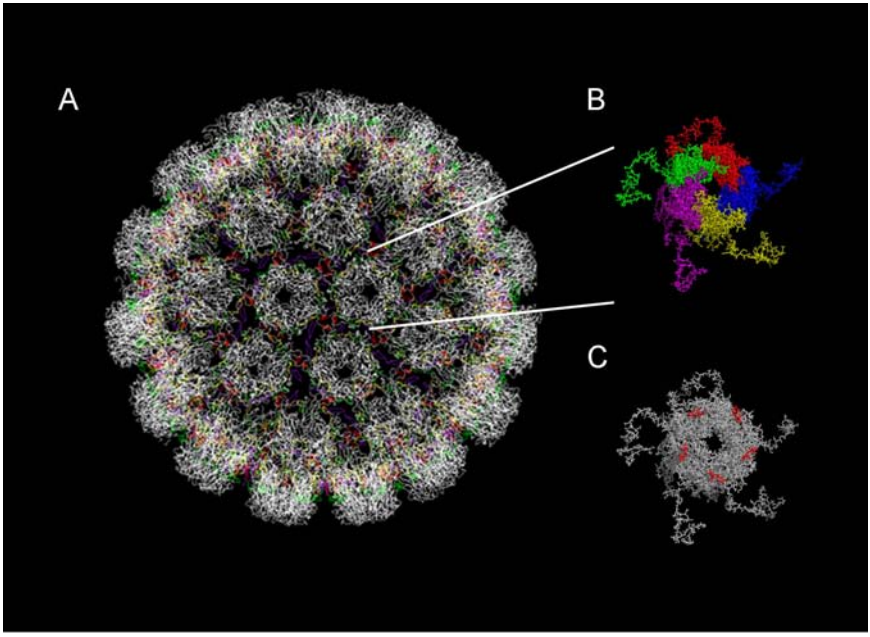


Fig. 1. Structures of the SV40 virion, SV40 VP1-pentamer, and sialyloligosaccharide receptor fragments binding to murine polyomavirus VP1-pentamer. (**A**) C-α backbone drawing of SV40 virion reconstructed from the structural unit of the SV40 VP1 capsid (1SVA) using VIPER. (**B**) Each SV40 VP1 protein in one VP1-pentamer is highlighted with different colors. (**C**) sialyloligosaccharide receptor fragments binding to murine polyomavirus VP1-pentamer (1SID) are highlighted with red.

different symmetry locations on the particle with intrinsic activity of the VP1. The capsid of SV40 also contains two internal proteins, VP2 and VP3, where VP3 is an amino terminally truncated form of VP2 (Fig. 3). Low resolution (25 Å) crystallographic studies of polyoma have shown that a part of VP2/3 inserts into the inward-facing cavity along the 5-fold axis of a VP1-pentamer.[5] *In vitro* binding studies have confirmed that the VP1-pentamer interacts tightly with VP2/3; that a sequence near the common C-terminus of VP2/3 is necessary and sufficient for complex formation[6]; and that VP3 residues 155 to 190 bind to VP1-pentamers.[7] Thus, the C-terminus of VP2/3 polypeptide chain binds in one of five equally probable orientations to its pentameric

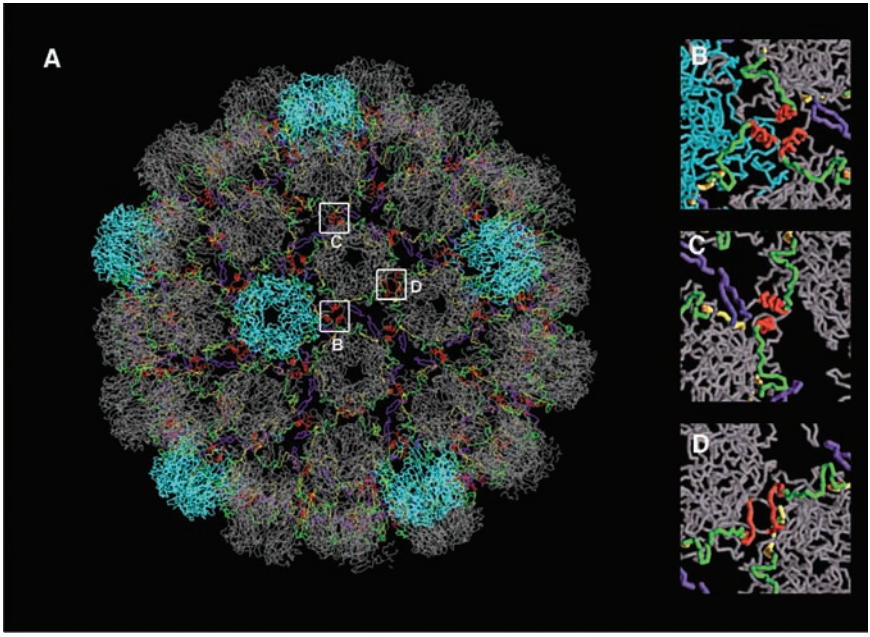


Fig. 2. Structures of the carboxyl-terminal regions in VP1-pentamer. (**A**) C-α backbone drawing of SV40 virion reconstructed from the structural unit of the SV40 VP1 capsid (1SVA) using VIPER. Pentavalent VP1-pentamers (VP1-pentamers surrounded by five VP1-pentamers) are marked in light blue. The C-terminal arm regions, VP1(301–312), (313–329), (330–344) and (345–361), are colored in red, green, yellow, and purple, respectively. The positions of three distinct interaction-types of the C-terminal arm in the SV40 particle are shown in (**B**) (α–α'–α''), (**C**) (β–β'), and (**D**) (γ–γ).

VP1-partner. X-ray crystallographic data of the C-terminal portion of the polyomavirus VP2 (residues 269–296) showed that the C-terminus of VP2/3 associates with VP1-pentamer tightly and specifically through hydrophobic interactions.[8] The larger N-terminal part of the internal protein is flexible and sensitive to gentle proteolysis. At the center of the virion, it locates the virus genome as minichromosome arranged with histones electron density at the center of the SV40 is smeared, suggesting that the minichromosome is not organized with icosahedral symmetry matching the capsid symmetry (Fig. 3).[9]

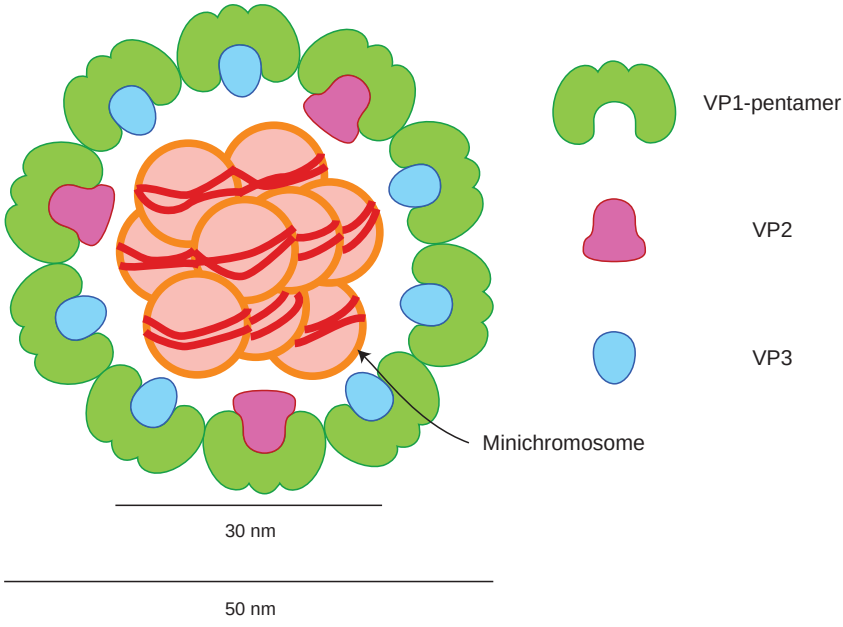


Fig. 3. Components of the SV40 virion. SV40 minor capsid protein VP2 or VP3 bind to one VP1-pentamer from inside the particle. SV40 virus genome arranges minichromosome with histone and is located at the center of the particle. Minichromosome inside the particle is thought to take on a spherical, 30 nm structure in diameter, without histone HI molecule.

Virion Assembly

One VP1-pentamer associates with one VP2 or the amino terminally truncated form VP3 soon after their synthesis in the cytoplasm (Fig. 4). It is because both capsid proteins harbor resident nuclear localization signals (NLSs),[10,11] that an NLS-defective SV40 VP3 can localize in the nucleus if wild-type VP1 with a functional NLS is coexpressed in the same cell, suggesting that VP1 and VP3 form a complex in the cytoplasm prior to nuclear entry. Also, regardless of the amount of capsid protein synthesized in the cytoplasm, a constant ratio of VP1 and VP3 that reflects the ratio in the mature virion is found in the nucleus,[12] implying that the stoichiometry of VP1 and

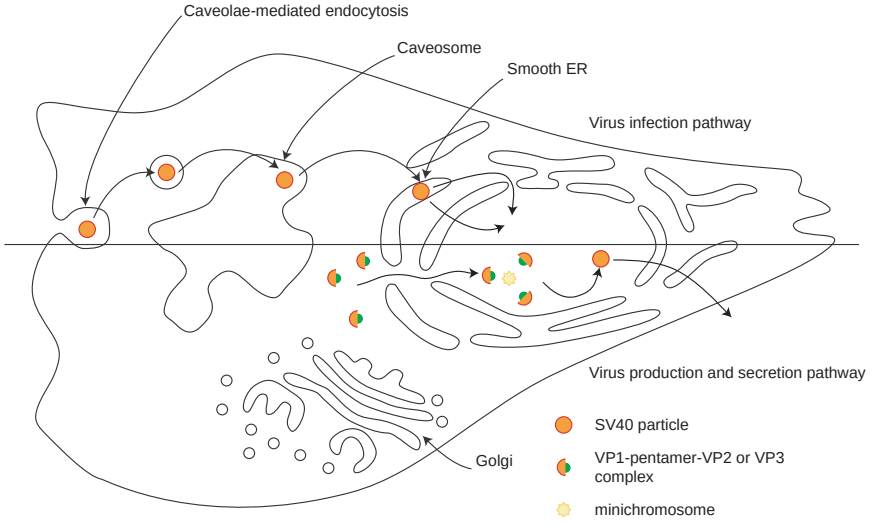


Fig. 4. Schematic representation of SV40 virus infection and secretion pathway. *Upper figure*. 1. SV40 virus first binds to the cell surface. 2. SV40 virus is endocytosed in caveolin 1 containing vesicles and delivered to caveosomes. 3. SV40 virus leaves for smooth endoplasmic reticulum. 4. SV40 viral proteins and host proteins may transport virus genome called minichromosme into nucleus. *Lower figure,* 1. SV40 VP1-pentamer binds to one minor capsid protein VP2 or VP3 in cytoplasm. 2. VP1-pentamer-VP2 or -VP3 complex transport into nucleus. 3. VP1-pentamer-VP2 or -VP3 complex select virus genome called minichromosome from the replicating and transcribing pools to be encapsidated. 4. Complete SV40 virion escapes from host cell with or without cell lysis.

VP3 in the nucleus is predetermined by formation of the complex in the cytoplasm. After formation of the VP1-pentamer-VP2 or -VP3 complex, the latter enters into the nucleus to enclose viral genome that complex with histone, termed minichromosome for the SV40 capsid assembly. From the isolated SV40 minichromosome structural analysis, it has been shown that minichromosome contains all five histones, including histone HI that exists in solution under approximately physiological ionic conditions as a compact roughly spherical particle approximately 300 Å in diameter in nucleus.[13] The *in vitro*

binding studies have suggested that all three capsid proteins bind DNA nonspecifically, raising the dilemma of how they attain specificity to the SV40 minichromosome in the presence of a large excess of genomic DNA.[14–18] VP3 has been shown to repress transcription from the viral early promoter *in vitro*, and that it cooperated with Sp1 through the Sp1 binding site on viral genome.[19–21] Therefore, Sp1 might recruit the VP1-pentamer-VP2/3 complex to minichromosome, conferring upon them specific DNA recognition. After association of VP1-pentamer to minichromosome, the carboxy terminus of VP1 responsible for the non-equivalent association of capsomeres interacts with the adjacent VP1-pentamer to form SV40 capsid.[22] During viral infection, the 72 KDa cellular chaperone heat shock cognate protein (hsc70) binds VP1 post-translation and colocalizes with VP1 to the nucleus.[23] Recombinant VP1 C-terminal domain from *E. coli* was copurified with the prokaryotic hsp70 chaperone DnaK. A member of the 70 kDa family of the cellular stress proteins assist in protein folding by preventing inappropriate intra- and intermolecular interactions during normal protein synthesis and transport and when the cells are exposed to a variety of environmental stresses. This suggests a role of the 70 kDa heat shock protein (hsp70) for the chaperones to regulate the quality and location of capsid assembly. In particle formation *in vitro*, particle assembly was promoted by the eukaryotic hsc70 protein, in combination with the J-domain function of the SV40 large T-antigen protein.[24] Thus, polyomavirus capsid assembly can be recapitulated with high-fidelity *in vitro*, using either prokaryotic or eukaryotic hsp70 chaperone systems, thereby supporting a role of cellular chaperones in the *in vivo* regulation of virion assembly. It also revealed that VP2 allowed particle assembly of VP1-pentamers into spherical particles in a pH range of 7.0 to 4.0 *in vitro*.[25] A region common to VP2 and VP3 (amino acids 119–272) was required to promote the VP1-pentamer assembly. These results are relevant to the control of recombinant capsid formation by minor capsid proteins. After particle formation, virions escape from the infected cells with cell lysis[26–30] or without cell lysis from apical cell surface.[31]

Host Cell Recognition and Entry

Class I major histocompatibility complex proteins have appeared to be the major cell surface receptors for SV40 (Fig. 5).[32,33] The class I molecules encoded by the major histocompatibility complex (MHC)

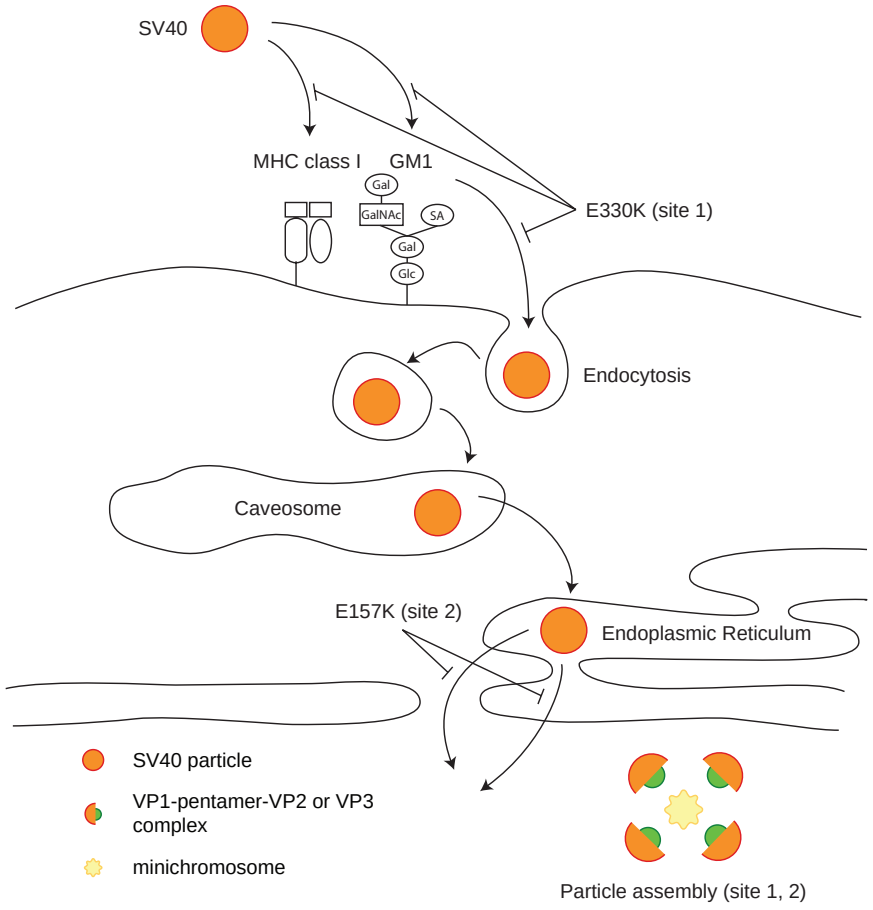


Fig. 5. Schematic reproduction of SV40 calcium binding site mutant particle phenotype in virus infectious pathway. E330K site 1 calcium binding site mutant is deficient in the host cell attachment and endocytosis. Another E157K site 2 calcium binding site mutant is deficient in virus genome transfer into nucleus inside host cell.

present endogenously synthesized antigenic peptide fragments to cytotoxic T lymphocytes. SV40 binding to cells can be blocked by two monoclonal antibodies against class I human lymphocyte antigen (HLA) proteins but not by monoclonal antibodies specific for other cell surface proteins. Also, SV40 does not bind to cells of two different human lymphoblastoid cell lines which do not express surface class I MHC proteins because of genetic defects in the β2-microglobulin gene in one line and in the HLA complex in the other. Transfection of these cell lines with cloned genes for β2-microglobulin and HLA-B8, respectively, restored expression of their surface class I MHC proteins and resulted in concomitant SV40 binding. Finally, SV40 binds to purified HLA proteins *in vitro* and selectively binds to class I MHC proteins in a surface extract. These proteins are an essential component of the cell surface receptor for SV40. For the polyomavirus receptor molecules, it has been shown that infectivity of the polyomavirus A2 strain in mouse Swiss 3T3 fibroblasts, is significantly reduced only in the presence of natural integrin ligands carrying an LDV motif or antibodies directed against the alpha4 and beta1 integrin subunit in the alpha4-deficient BALB/c 3T3 cells, increases viral infectivity.[34] Addition of alpha4 function-blocking antibodies, prior to or after virus adsorption, blocks this increased infectivity without affecting virus binding to cells. These data indicate that expression of alpha4 integrin enhances permissivity to polyomavirus, probably by acting as one of the post-attachment receptors.

The important role of sialic acid in the adsorption of polyomavirus to cells was first demonstrated by the abolition of viral-mediated hemagglutination upon treatment of erythorocytes with neuraminidase.[35–37] Neuraminidase has also been used successfully to prevent polyomavirus infection of cultured mouse embryo cells. These observations provide evidence for sialic acid as an integral component of the cell surface receptors for polyoma virus. The crystal structure of a recombinant polyomavirus VP1 pentamer (residues 32–320), in complex with a branched disialylated hexasaccharide receptor fragment, extends our understanding of oligosaccharide receptor recognition (Fig. 1C).[38,39] The model presented here offers a much more refined view of the interactions that determine carbohydrate

recognition and allows us to assign additional specific contacts, in particular those involving the (α2,6)-linked, branching sialic acid. The structure of the unliganded VP1-pentamer, determined independently, shows that the oligosaccharide fits into a preformed groove and induces no measurable structural rearrangements. A comparison with assembled VP1 in the virus capsid reveals a rearrangemnt of residues 32–45 at the base of the pentamer. This segment may help prevent the formation of incorrectly assembled particles by reducing the likelihood that the C-terminal arm will fold back into its pentamer of origin. Biochemical assay improved the notion for polyomavirus binding to oligosaccharide receptor recognition. Binding and flotation assays showed that addition of gangliosides to the phospholipids vesicles allowed specific binding of the respective viruses. Specific gangliosides can serve as plasma membrane receptors, GD1a and GT1b for polyomavirus, GM1 for SV40,[40] and GD1b and GT1b for BK.[41]

After host cell surface binding through cell surface proteins and gangliosides, it has been suggested that the local structure of virion changes during endocytosis[42] and this change is involved with its calcium binding sites.[43,44] Several lines of evidence have suggested that this inter VP1-pentamer interaction is strengthened by calcium ion chelation and disulfide bonding. Structural refinement on SV40, in which divalent calcium ions were replaced with trivalent gadolinium ions, has identified two probable sites, termed site-1 and site-2, of calcium ion coordination per VP1 monomer on the capsid (Fig. 5). Site-1 consists of the Glu 216 side chain and Ser 213 of one VP1 monomer, the Glu 46 and Glu 48 side chains of a second VP1 monomer from the same pentamer, and the Glu 330 side chain (C-terminal arm) of a third VP1 monomer from a neighboring pentamer. Site-2 consists of the Glu 157, Glu 160, and Glu 216 side chains and Lys 214 carbonyl oxygen of the first VP1 monomer and the Asp 345 side chain (C-terminal arm) of the third monomer. Each pair of calcium ions is expected to tie together two different VP1 pentamers by interacting with mostly acidic amino acid residues contributed by three VP1 chains. The point mutant E330K VLPs of SV40, site-1 calcium mutant, turned out to be defective in adsorbing to cells.[43] Conceivably, the salt bridges in place of calcium ion at site-1 could have prevented

the E330K VLP from undergoing the calcium-dependent structural shift necessary for binding to the cell receptor, the major histocompatibility complex (MHC) class I molecule, leading to the adsorption defect. Analyzing the structural difference between the mutant E330K VLP and the wild-type virion could help delineate the virus site or epitope responsible for attachment to MHC class I. Another point mutant E157K, the site-2 calcium mutant, about 10% larger in diameter than the wild type, was able to enter cells, but this did not lead to T-antigen expression.[44] Cell-internalized E157K DNA effectively coimmunoprecipitated with anti-VP1 antibody, but little of the DNA did so with anti-VP3 antibody, and none was detected in anti-importins immunoprecipitate. Yet a substantial amount of VP3 was present in anti-VP1 immune complexes, suggesting that internalized E157K particles are ineffective at exposing VP3. These results show that mutant E157K infection is blocked at a stage prior to the interaction of VP3 nuclear localization signal with importins, consistent with a role for calcium-binding site-2 in post-entry steps leading to the nuclear import of the infecting SV40.

Virion Disassembly and Viral Genome Entrance into the Nucleus

SV40 is unusual among animal viruses in that it enters cells through caveolae, and the internalized virus accumulates in a smooth endoplasmic reticulum (ER) compartment (Figs. 4 and 5). After associating with caveolae, SV40 leaves the plasma membrane in small, caveolin-1-containing vesicles.[45–47] It then enters larger, peripheral organelles with a non-acidic pH. Although rich in caveolin-1, these organelles do not contain markers for endosomes, lysosomes, ER or Golgi, nor do they acquire ligands of clathrin-coated vesicle endocytosis. After several hours in these organelles, SV40 is sorted into tubular, caveolin-free membrane vesicles that move rapidly along microtubules, and is deposited in perinuclear, syntaxin 17-positive, smooth ER organelles. There exists a two-step transport pathway from the plasma-membrane caveolar, through an intermediate organelle termed the caveosome, to the ER. This pathway bypasses endosomes and the

Golgi complex, and is part of the productive infectious route used by SV40. There is no SV40 in the cell nucleus after a 16-hour infection.[48] It is assumed that SV40 in ER dissociated into VP1-pentamers to release minichromosome.

There has been suggested that treatment of polyoma virions with EGTA and dithiothreitol (DTT) resulted in the dissociation of the virions into a DNA-protein complex and individual structural capsomere subunits *in vitro*.[49,50] It is interesting to presume that this artificial dissociation partially reflect the actual biological event of SV40 in ER. Recently, a protein disulfide isomerase (PDI)-like protein, ERp29, exposes the C-terminal arm of polyomavirus VP1 protein, leading to formation of a hydrophobic particle that binds to a lipid bilayer; this reaction likely mimics initiation of polyomavirus penetration across the ER membrane.[51] These results thus identify an ER factor that mediates membrane penetration of a nonenveloped virus and suggest that PDI family members are generally involved in ER remodeling reactions. Although not essential, the reducing agent DTT and the calcium-chelator EGTA stimulated the ER-induced polyomavirus conformational change. It is possible that DTT and EGTA partially destabilized polyomavirus structure, enabling ERp29 to subsequently expose VP1's C-terminal arm efficiently. This explanation is supported by previous biochemical and X-ray structural studies that showed that disulfide bridges and calcium ions provide critical structural support for polyomavirus.[2,50] Because the disulfide bond in polyomavirus virions is in the proximity of the N terminus of VP1, its reduction may enable ERp29 to also expose the VP1 N terminus. DTT and EGTA likely mimicked the action of ER reductases (e.g. PDI) and calcium binding proteins (e.g. calnexin and calreticulin) that would normally act on the virus. It has been suggested that VP2 and VP3 are involved in the transport of minichromosome into the nucleus. It has been shown that VP2 and VP3 nuclear localization signals have important roles in nuclear transport of minichromosome with importins.[52] VP2 and VP3 NLS were partially exposed outside of the virion in ER, then after detachment of partially dissociated virion associates with importins and transport minichromosome into nucleus through nuclear pore complex. It has also been suggested that capsid disassembly within the ER

lumen, VP3 and perhaps VP2, oligomerize and integrate into the ER membrane or nuclear matrix, potentially creating a viroporin that aids in viral DNA transport.[53]

Encapsulation of Exogenous DNA, Chemicals, Proteins for Agent Delivery into Cells

SV40 is an attractive potential vector with high-efficiency gene transfer into a wide variety of human tissues, a critical target organ for the cure of many diseases (Fig. 6). It has been shown that the VP1 proteins of SV40[24,54–59] and closely related viruses such as JC,[60,61] BK,[62–64] and

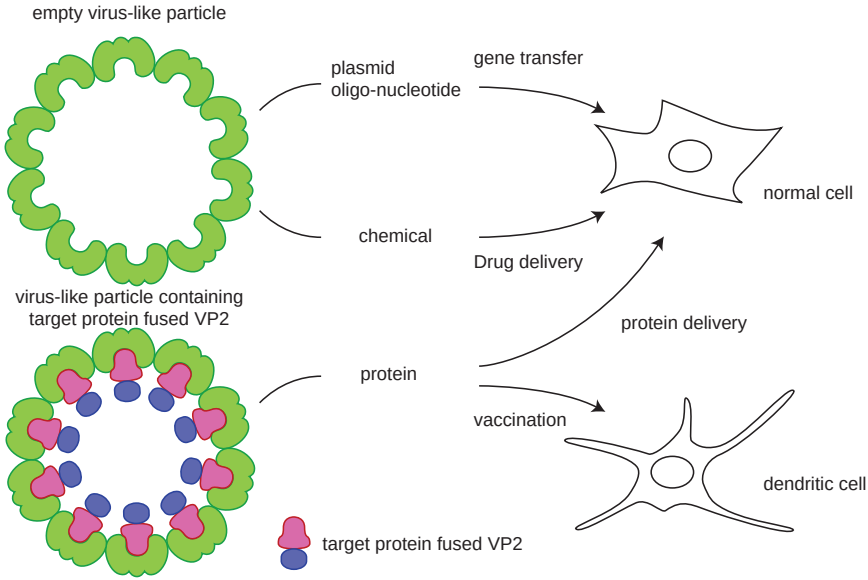


Fig. 6. Schematic reproduction of virus-like particle in polyomavirus for the use of exogenous genes, chemicals, and proteins. Empty virus-like particles can be used for gene and chemical delivery into normal cells. Exogenous proteins fused with minor capsid proteins can be packed inside the particle when co-expressed in insect cells or by using *in vitro* packaging systems. These exogenous packaged virus-like particles can be used for protein transfer into normal cells and for vaccination into dendritic cells.

murine polyomavirus[65–68] form virus-like particles (VLPs) when expressed in insect cells from baculoviral vectors. Electron and confocal microscopy have revealed the entry and the movement of polyomavirus virions and artificial virus-like particles (VP1 pseudocapsids) in mouse fibroblasts and epithelial cells.[69] There are no visible difference in adsorption and VP1 pseudocapsid ("empty" or containing DNA). Viral particles entered cells internalized in smooth monopinocytic vesicles, often in the proximity of larger, caveola-like invaginations. Both the "empty" vesicles derived from caveolae and the vesicles containing viral particles were stained with the anti-caveolin-1 antibody, and the two types of vesicles often fused in the cytoplasm. By 3 hour postinfection, a strong signal of the VP1 (but no viral particles) had accumulated in the proximity of the nuclei, around the outer nuclear membrane. The vast majority of VP1 pseudocapsids did not enter the nuclei as seen in SV40 virions.[48] These results provide important information for the use of VLPs as a tool of drug delivery carrier. We can assume that VLPs can enter into cells in the same way same as virions. Nuclear extracts containing the three proteins of SV40 (VP1, VP2, and VP3) were allowed to interact with purified SV40 DNA, or with plasmid DNA.[59] These mixtures can deliver DNAs into the cells. Polyomavirus VP1 capsid particles expressed in *E. coli* and Sf9 were also packed *in vitro* with anti-sense oligonucleotides[70] and plasmid DNA.[71] Furthermore, JC virus pseudocapsids prepared from *E. coli* were also able to deliver exogenous DNA into human fetal kidney epithelial cells.[72] These results indicate that recombinant JC virus VP1 is able to self-assemble into capsid-like particles and to package DNA in the absence of the minor capsid proteins, VP2 and VP3. These prokaryotic assembly systems may facilitate the investigation of maturation mechanism(s) of polyomaviruses. Furthermore, capsid-like particles of these VP1 proteins in polyomaviridae could potentially be used as a human gene transfer vector.

Not only plasmid DNA but also encapsulation of chemicals was accomplished with human polyomavirus JC virus (JCV).[73] Propidium iodide (PI) was packaged into VP1–VLP as a reporter molecule. The fluorescence intensity of the VP1–VLP depended strongly on the initial PI concentration. This packaging method is easy to handle and is

applicable to viruses and VP1–VLP which can be dissociated and reassociated chemically. It is potentially useful for the introduction of encapsulated chemicals into ER.

Exogenous protein encapsulation into VLPs was accomplished by using the VP1 binding motif of minor capsid proteins. Co-expression of EGFP fused with the VP1 binding motif of minor capsid proteins and VP1 in Sf9 cells produce EGFP containing VLPs. EGFP-VLPs were analyzed for their ability to be internalized and processed by mouse cells and to activate mouse and human dendritic cells (DC) *in vitro*.[74] EGFP–VLPs entered the mouse epithelial cells, fibroblasts and human and mouse DC efficiently and were processed by both, as well as lysosomes and proteasomes. These results provide the basis for the preparation of mouse polyomavirus capsid-like particles for transfer of foreign peptides or proteins into cells. The use of foreign peptide fused with VP2 containing VLPs for vaccination has been shown. Murine polyomavirus (MPyV), VLPs, containing a fusion protein between MPyV VP2 and the extracellular and transmembrane domain of HER-2/neu (Her2),[75] Her2 1-683 PyVLPs, were tested for their ability to vaccinate against Her2-expressing tumors in two different *in vivo* models. Protection was assessed both against a lethal challenge with a BALB/c mammary carcinoma transfected with human Her2 (D2F2/E2) and against the outgrowth of autochthonous mammary carcinomas in BALB-neuT mice, transgenic for the activated rat Her2 oncogene. A single injection of Her2 1-683 PyVLPs before tumor inoculation induced a complete rejection of D2F2/E2 tumor cells in BALB/c mice. Similarly, a single injection of Her2 1-683 PyVLPs at six weeks of age protected the BALB/neuT mice with atypical hyperplasia from a later outgrowth of mammary carcinomas; all the controls developed palpable tumors in all their mammary glands. VLPs containing only VP1 and VP2 did not induce protection. The protection elicited by Her2 1-683 PyVLPs vaccination was most likely due to a cellular immune response, whereas antibodies against Her2 were not detected in both the two models. The results show it is feasible to use MPyV-VLPs carrying Her2 fusion proteins as safe and efficient vaccines against Her2-expressing tumors.

Acknowledgments

This work was supported in part by Grants from the Uehara Memorial Foundation.

References

1. Liddington RC, Yan Y, Moulai J, *et al.* (1991) Structure of simian virus 40 at 3.8-A resolution. *Nature* **354**: 278–284.
2. Stehle T, Gamblin SJ, Yan Y, Harrison SC. (1996) The structure of simian virus 40 refined at 3.1 A resolution. *Structure* **4**: 165–182.
3. Yan Y, Stehle T, Liddington RC, *et al.* (1996) Structure determination of simian virus 40 and murine polyomavirus by a combination of 30-fold and 5-fold electron-density averaging. *Structure* **4**: 157–164.
4. Salunke DM, Caspar DL, Garcea RL. (1986) Self-assembly of purified polyomavirus capsid protein VP1. *Cell* **46**: 895–904.
5. Griffith JP, Griffith DL, Rayment I, *et al.* (1992) Inside polyomavirus at 25-A resolution. *Nature* **355**: 652–654.
6. Barouch DH, Harrison SC. (1994) Interactions among the major and minor coat proteins of polyomavirus. *J Virol* **68**: 3982–3989.
7. Nakanishi A, Nakamura A, Liddington R, Kasamatsu H. (2006) Identification of amino acid residues within simian virus 40 capsid proteins Vp1, Vp2, and Vp3 that are required for their interaction and for viral infection. *J Virol* **80**: 8891–8898.
8. Chen XS, Stehle T, Harrison SC. (1998) Interaction of polyomavirus internal protein VP2 with the major capsid protein VP1 and implications for participation of VP2 in viral entry. *EMBO J* **17**: 3233–3240.
9. Baker TS, Drak J, Bina M. (1988) Reconstruction of the three-dimensional structure of simian virus 40 and visualization of the chromatin core. *Proc Natl Acad Sci USA* **85**: 422–426.
10. Clever J, Kasamatsu H. (1991) Simian virus 40 Vp2/3 small structural proteins harbor their own nuclear transport signal. *Virology* **181**: 78–90.
11. Ishii N, Nakanishi A, Yamada M, *et al.* (1994) Functional complementation of nuclear targeting-defective mutants of simian virus 40 structural proteins. *J Virol* **68**: 8209–8216.
12. Lin W, Hata T, Kasamatsu H. (1984) Subcellular distribution of viral structural proteins during simian virus 40 infection. *J Virol* **50**: 363–371.
13. Varshavsky AJ, Nedospasov SA, Schmatchenko VV, *et al.* (1977) Compact form of SV40 viral minichromosome is resistant to

nuclease: possible implications for chromatin structure. *Nucl Acids Res* **4**: 3303–3325.
14. Li PP, Nakanishi A, Shum D, *et al.* (2001) Simian virus 40 Vp1 DNA-binding domain is functionally separable from the overlapping nuclear localization signal and is required for effective virion formation and full viability. *J Virol* **75**: 7321–7329.
15. Soussi T. (1986) DNA-binding properties of the major structural protein of simian virus 40. *J Virol* **59**: 740–742.
16. Moreland RB, Montross L, Garcea RL. (1991) Characterization of the DNA-binding properties of the polyomavirus capsid protein VP1. *J Virol* **65**: 1168–1176.
17. Chang D, Cai X, Consigli RA. (1993) Characterization of the DNA binding properties of polyomavirus capsid protein. *J Virol* **67**: 6327–6331.
18. Clever J, Dean DA, Kasamatsu H. (1993) Identification of a DNA binding domain in simian virus 40 capsid proteins Vp2 and Vp3. *J Biol Chem* **268**: 20877–20883.
19. Gordon-Shaag A, Ben-Nun-Shaul O, Roitman V, *et al.* (2002) Cellular transcription factor Sp1 recruits simian virus 40 capsid proteins to the viral packaging signal, ses. *J Virol* **76**: 5915–5924.
20. Oppenheim A, Sandalon Z, Peleg A, *et al.* (1992) A cis-acting DNA signal for encapsidation of simian virus 40. *J Virol* **66**: 5320–5328.
21. Dalyot-Herman N, Ben-nun-Shaul O, Gordon-Shaag A, *et al.* (1996) The simian virus 40 packaging signal ses is composed of redundant DNA elements which are partly interchangeable. *J Mol Biol* **259**: 69–80.
22. Garcea RL, Salunke DM, Caspar DL. (1987) Site-directed mutation affecting polyomavirus capsid self-assembly *in vitro*. *Nature* **329**: 86–87.
23. Cripe TP, Delos SE, Estes PA, Garcea RL. (1995) *In vivo* and *in vitro* association of hsc70 with polyomavirus capsid proteins. *J Virol* **69**: 7807–7813.
24. Chromy LR, Pipas JM, Garcea RL. (2003) Chaperone-mediated *in vitro* assembly of Polyomavirus capsids. *Proc Natl Acad Sci USA* **100**: 10477–10482.
25. Kawano MA, Inoue T, Tsukamoto H, *et al.* (2006) The VP2/VP3 minor capsid protein of simian virus 40 promotes the *in vitro* assembly of the major capsid protein VP1 into particles. *J Biol Chem* **281**: 10164–10173.
26. Daniels R, Sadowicz D, Hebert DN. (2007) A Very Late Viral Protein Triggers the Lytic Release of SV40. *PLoS Pathog* **3**: e98.
27. Granboulan N, Tournier P, Wicker R, Bernhard W. (1963) An electron microscope study of the development of SV40 virus. *J Cell Biol* **17**: 423–441.

28. Oshiro LS, Rose HM, Morgan C, Hsu KC. (1967) Electron microscopic study of the development of simian virus 40 by use of ferritin-labeled antibodies. *J Virol* **1**: 384–399.
29. Maul GG, Rovera G, Vorbrodt A, Abramczuk J. (1976) Membrane fusion as a mechanism of simian virus 40 entry into different cellular compartments. *J Virol* **28**: 936–944.
30. Maul GG. (1976) Fibrils attached to the nuclear pore prevent egress of SV40 particles from the infected nucleus. *J Cell Biol* **70**: 714–719.
31. Clayson ET, Brando LV, Compans RW. (1989) Release of simian virus 40 virions from epithelial cells is polarized and occurs without cell lysis. *J Virol* **63**: 2278–2288.
32. Atwood WJ, Norkin LC. (1989) Class I major histocompatibility proteins as cell surface receptors for simian virus 40. *J Virol* **63**: 4474–4477.
33. Breau WC, Atwood WJ, Norkin LC. (1992) Class I major histocompatibility proteins are an essential component of the simian virus 40 receptor. *J Virol* **66**: 2037–2045.
34. Caruso M, Belloni L, Sthandier O, *et al.* (2003) Alpha4beta1 integrin acts as a cell receptor for murine polyomavirus at the postattachment level. *J Virol* 77: 3913–3921.
35. Eddy BE, Rowe WP, Hartley JW, *et al.* (1958) Hemagglutination with the SE polyoma virus. *Virology* **6**: 290–291.
36. Fried H, Cahan LD, Paulson JC. (1981) Polyoma virus recognizes specific sialyligosaccharide receptors on host cells. *Virology* **109**: 188–192.
37. Cahan LD, Singh R, Paulson JC. (1983) Sialyloligosaccharide receptors of binding variants of polyoma virus. *Virology* **130**: 281–289.
38. Stehle T, Harrison SC. (1996) Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. *Structure* **4**: 183–194.
39. Stehle T, Harrison SC. (1997) High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. *EMBO J* **16**: 5139–5148.
40. Tsai B, Gilbert JM, Stehle T, *et al.* (2003) Gangliosides are receptors for murine polyoma virus and SV40. *EMBO J*, **22**: 4346–4355.
41. Low JA, Magnuson B, Tsai B, Imperiale MJ. (2006) Identification of gangliosides GD1b and GT1b as receptors for BK virus. *J Virol* **80**: 1361–1366.
42. Cavaldesi M, Caruso M, Sthandier O, *et al.* (2004) Conformational changes of murine polyomavirus capsid proteins induced by sialic acid binding. *J Biol Chem* **279**: 41573–41579.

43. Li PP, Naknanishi A, Tran MA, *et al.* (2003) Importance of Vp1 calcium-binding residues in assembly, cell entry, and nuclear entry of simian virus 40. *J Virol* **77**: 7527–7538.
44. Li PP, Nguyen AP, Qu Q, *et al.* (2007) Importance of calcium-binding site 2 in simian virus 40 infection. *J Virol* **81**: 6099–6105.
45. Pelkmans L, Kartenbeck J, Helenius A. (2001) Caveolar endocytosis of simian virus 40 reveals a new two-step vesicular-transport pathway to the ER. *Nat Cell Biol* **3**: 473–483.
46. Norkin LC, Anderson HA, Wolfrom SA, Oppenheim A. (2002) Caveolar endocytosis of simian virus 40 is followed by brefeldin A-sensitive transport to the endoplasmic reticulum, where the virus disassembles. *J Virol* **76**: 5156–5166.
47. Richards AA, Stang E, Pepperkok R, Parton RG. (2002) Inhibitors of COP-mediated transport and cholera toxin action inhibit simian virus 40 infection. *Mol Biol Cell* **13**: 1750–1764.
48. Kartenbeck J, Stukenbrok H, Helenius A. (1989) Endocytosis of simian virus 40 into the endoplasmic reticulum. *J Cell Biol* **109**: 2721–2729.
49. Brady JN, Winston VD, Consigli RA. (1978) Dissociation of polyoma virus by the chelation of calcium ions found associated with purified virions. *J Virol* **23**: 717–724.
50. Brady JN, Winston VD, Consigli RA. (1978) Characterization of a DNA-protein complex and capsomere subunits derived from polyoma virus by treatment with ethyleneglycol-bis-N,N′-tetraacetic acid and dithiothreitol. *J Virol* **27**: 193–204.
51. Magnuson B, Rainey EK, Benjamin T, *et al.* (2005) ERp29 triggers a conformational change in polyomavirus to stimulate membrane binding. *Mol Cell* **20**: 289–300.
52. Nakanishi A, Shum D, Morioka H, *et al.* (2002) Interaction of the Vp3 nuclear localization signal with the importin alpha 2/beta heterodimer directs nuclear entry of infecting simian virus 40. *J Virol* **76**: 9368–9377.
53. Daniels R, Rusan NM, Wadsworth P, Hebert DN. (2006) SV40 VP2 and VP3 insertion into ER membranes is controlled by the capsid protein VP1: implications for DNA translocation out of the ER. *Mol Cell* **24**: 955–966.
54. Ishizu KI, Watanabe H, Han SI, *et al.* (2001) Roles of disulfide linkage and calcium ion-mediated interactions in assembly and disassembly of virus-like particles composed of simian virus 40 VP1 capsid protein. *J Virol* **75**: 61–72.

55. Kanesashi SN, Ishizu K, Kawano MA, *et al.* (2003) Simian virus 40 VP1 capsid protein forms polymorphic assemblies *in vitro*. *J Gen Virol* **84**: 1899–1905.
56. Kosukegawa A, Arisaka F, Takayama M, *et al.* (1996) Purification and characterization of virus-like particles and pentamers produced by the expression of SV40 capsid proteins in insect cells. *Biochim Biophys Acta* **1290**: 37–45.
57. Colomar MC, Degoumois-Sahli C, Beard P. (1993) Opening and refolding of simian virus 40 and *in vitro* packaging of foreign DNA. *J Virol* **67**: 2779–2786.
58. Kimchi-Sarfaty C, Arora M, Sandalon Z, *et al.* (2003) High cloning capacity of *in vitro* packaged SV40 vectors with no SV40 virus sequences. *Hum Gene Ther* **14**: 167–177.
59. Sandalon Z, Dalyot-Herman N, Oppenheim AB, Oppenheim A. (1997) *In vitro* assembly of SV40 virions and pseudovirions: vector development for gene therapy. *Hum Gene Ther* **8**: 843–849.
60. Chang D, Fung CY, Ou WC, *et al.* (1977) Self-assembly of the JC virus major capsid protein, VP1, expressed in insect cells. *J Gen Virol* **78** (Pt 6): 1435–1439.
61. Goldmann C, Petry H, Frye S, *et al.* (1999) Molecular cloning and expression of major structural protein VP1 of the human polyomavirus JC virus: formation of virus-like particles useful for immunological and therapeutic studies. *J Virol* **73**: 4465–4469.
62. Touze A, Bousarghin L, Ster C, *et al.* (2001) Gene transfer using human polyomavirus BK virus-like particles expressed in insect cells. *J Gen Virol* **82**: 3005–3009.
63. Viscidi RP, Rollison DE, Viscidi E, *et al.* (2003) Serological cross-reactivities between antibodies to simian virus 40, BK virus, and JC virus assessed by virus-like-particle-based enzyme immunoassays. *Clin Diagn Lab Immunol* **10**: 278–285.
64. Li TC, Takeda N, Kato K, *et al.* (2003) Characterization of self-assembled virus-like particles of human polyomavirus BK generated by recombinant baculoviruses. *Virology* **311**: 115–124.
65. Li M, Delos SE, Montross L, Garcea RL. (1995) Polyomavirus VP1 phosphorylation: coexpression with the VP2 capsid protein modulates VP1 phosphorylation in Sf9 insect cells. *Proc Natl Acad Sci USA* **92**: 5992–5996.

66. Pawlita M, Muller M, Oppenlander M, *et al.* (1996) DNA encapsidation by viruslike particles assembled in insect cells from the major capsid protein VP1 of B-lymphotropic papovavirus. *J Virol* **70**: 7517–7526.
67. Forstova J, Krauzewicz N, Wallace S, *et al.* (1993) Cooperation of structural proteins during late events in the life cycle of polyomavirus. *J Virol* **67**: 1405–1413.
68. Gillock ET, Rottinghaus S, Chang D, *et al.* (1977) Polyomavirus major capsid protein VP1 is capable of packaging cellular DNA when expressed in the baculovirus system. *J Virol* **71**: 2857–2865.
69. Richterova Z, Liebl D, Horak M, *et al.* (2001) Caveolae are involved in the trafficking of mouse polyomavirus virions and artificial VP1 pseudocapsids toward cell nuclei. *J Virol* **75**: 10880–10891.
70. Braun H, Boller K, Lower J, *et al.* (1999) Oligonucleotide and plasmid DNA packaging into polyoma VP1 virus-like particles expressed in *Escherichia coli*. *Biotechnol Appl Biochem* **29** (Pt 1): 31–43.
71. Stokrova J, Palkova Z, Fischer L, *et al.* (1999) Interactions of heterologous DNA with polyomavirus major structural protein, VP1. *FEBS Lett* **445**: 119–125.
72. Ou WC, Wang M, Fung CY, *et al.* (1999) The major capsid protein, VP1, of human JC virus expressed in *Escherichia coli* is able to self-assemble into a capsid-like particle and deliver exogenous DNA into human kidney cells. *J Gen Virol* **80** (Pt 1): 39–46.
73. Goldmann C, Stolte N, Nisslein T, *et al.* (2000) Packaging of small molecules into VP1-virus-like particles of the human polyomavirus JC virus. *J Virol Methods* **90**: 85–90.
74. Boura E, Liebl D, Spisek R, *et al.* (2005) Polyomavirus EGFP-pseudocapsids: analysis of model particles for introduction of proteins and peptides into mammalian cells. *FEBS Lett* **579**: 6549–6558.
75. Tegerstedt K, Lindencrona JA, Curcio C, *et al.* (2005) A single vaccination with polyomavirus VP1/VP2Her2 virus-like particles prevents outgrowth of HER-2/neu-expressing tumors. *Cancer Res* **65**: 5953–5957.
76. Adolph KW, Caspar DL, Hollingshead CJ, *et al.* (1979) Polyoma virion and capsid crystal structures. *Science* **203**: 1117–1120.
77. Rayment I, Baker TS, Caspar DL, Murakami WT. (1982) Polyoma virus capsid structure at 22.5 A resolution. *Nature* **295**: 110–115.
78. Nilsson J, Miyazaki N, Xing L, *et al.* (2005) Structure and assembly of a T = 1 virus-like particle in BK polyomavirus. *J Virol* **79**: 5337–5345.

This page intentionally left blank

Index

3D electron density map 562
3D-structure of the N protein 558
3D-structure 553, 556, 558, 567
3′-nontranslated region 152–154
5′-nontranslated region 152–154, 161
6 M urea 560

ACE2 81–84
alfalfa mosaic virus 174–176, 180–183, 190
amino acid 201, 202, 209, 211, 213, 215–220, 222, 223, 226–228, 230–232, 234, 235, 237, 238, 243
amino acid representation 228
ANN weights 223
analysis tools 267, 272, 275, 277, 282–284, 286
anti-parallel coiled-coil 557
antibodies 598, 599, 601, 603
antigen processing 602
antigenic 204–209, 223, 229
apoptosis 93, 96, 100, 103–105, 112, 113, 571, 577, 584, 589, 590
artificial amino acid 238
artificial neural network (ANN) 223, 225–233, 235–238, 242, 246
assembly 331, 332, 463, 464, 466, 472, 475, 479
assembly, concerted 174
assembly, sequential 174
asymmetric dimer 468, 469
atomic force microscopy 289–300, 302–306
Avastrovirus 572

B-cell epitope 203
bacteriophage 174, 176, 177
bacteriophage phi29 436
baculovirus-driven system 567
Bayesian network (BN) 212, 243–246
beta1 integrins 554
beta3 554
betaherpesvirus 485, 486, 496, 506
BHK-21 572, 585
binomial distribution 406, 409, 412, 418, 432, 433
bioinformatics 267–271
Bioinformatic Resource Center (BRC) 268–273
BK 609, 610, 618, 621
budding 555, 558
Bunyaviridae 553

CaCo-2 572, 573, 577, 584–586, 590

calicivirus 272–274, 279
cap-snatching mechanism 554
capsid 382, 383, 386–390, 393
capsid assembly 486, 489, 491, 496, 502, 508
capsid protein 173, 174, 176, 183
caveolae 609, 614, 619, 622
CEACAM1 81, 83–85
cell survival 93, 96, 108, 109, 111–113
CHIMERA 561
clathrin-mediated pathway 554
coat protein 176–183
codon-modification 520, 523
coiled-coil helices 559, 561, 564
common multiple of two and three 432, 434
connector 406, 410, 411, 433, 434, 436
core 153, 154
coronavirus 77–81, 84, 85, 272–274, 278, 279
cryo-electron tomography (cryoET) 493–495, 499–501
cryo-EM 359
cryo-EM analysis 483, 489, 493, 558
C-terminal alpha-helices 559
C-terminal domains 557
CTL epitopes 229, 230, 232
curve-shaped monomers 558
cyberinfrastructure 268, 269, 282
cytomegalovirus 485, 486, 506
cytoplasmic tail of Gn protein 554

database 268–272, 274, 277, 282, 284
data post-processing 215, 220
data pre-processing 215, 219
Daxx-pathway members 556
de novo design 212, 233–235, 237, 246
death domain 582, 584, 586, 587
density maps of N-protein 562
design 200–207, 209–213, 219, 223, 224, 231–239, 246
detergent-resistant membrane 158, 159
dimerization 189, 190
directed evolution (DE) 201, 207, 209, 210
DnaK 609, 615
DNA packaging 491, 499, 506
DNA packaging motor 406, 436
DNA vaccines 598–604
double-layered capsid 463, 464, 470, 472, 475
DTT 609, 620

E1 153, 154
E2 153, 154
Ebola virus 47, 52, 312–319
EGTA 609, 620
electron cryomicroscopy 449, 456, 462
electrostatic complementarity 475, 477, 479
EMAN 561, 566
encapsidate RNA 555, 564
encapsidation 176, 180, 183, 187
encoding 208, 215–217, 219, 220, 226, 227, 230–232, 240, 242
endocytosis 1, 2, 10, 11, 14, 20, 70, 609, 614, 616, 618, 619

endosome 64, 69, 70, 82–84
entry 126, 128, 129, 131, 135, 136, 137, 140
envelope proteins 349, 352, 362, 506
epitope 204–208, 223–225, 227, 229–236
ER 609, 614, 619–621, 623
evolutionary algorithm (EA) 228, 237, 239–242, 246
evolutionary search 237, 241, 242

feature representation 215, 219
fibrous RNP complexes 565
filamentous ribonucleoproteins (RNPs) 554
Flaviviridae 151, 156
fluid mosaic model 43, 45
formation of RNPs 556

gammaherpesvirus 484, 486, 488, 496, 497, 501, 502, 504, 506
gangliosides 609, 618
gastroenteritis 381
Gc proteins 555
gene transfer 609, 621, 622
genetic algorithm (GA) 239, 241, 242
genetic programming (GP) 242, 243
genome RNA (vRNA) 554
genomics 268, 276, 277
genus-specific mode 557
glycoprotein 312, 313, 316, 317, 330, 331, 346, 347, 351, 356
Gn-Gc heterodimers 554
Gn proteins 555
gp16 410, 411, 424
gp41 332–337, 340, 341, 347, 352, 354, 358, 360, 361
gp120 330, 331, 333–341, 345–348, 350–361
gp140 349
gp160 334, 349

hCD81 127, 129, 131, 134–139
HCV 125–140
HCV-LP 129–131, 136–138, 140
haemorrhagic fever with renal syndrome 554
Hantaviral minigenomes 557
hantaviral N protein oligomerization 557
hantaviral N-ptotein trimer 553, 556
hantaviral replication cycle 555
hantavirus 553–558, 564
hantavirus (cardio) pulmonary syndrome 554
hantavirus N protein 555
hantavirus replication cycle 554
helicase 151, 155, 161, 573, 574, 582, 583
hemagglutinin (HA) 48–54
hepatitis 91–98, 100, 101, 103, 105, 107, 109, 110, 113
hepatitis A 272–274, 280
hepatitis A virus (HAV) 200
hepatitis B virus (HBV) 199, 200, 206, 209
hepatitis C virus 151
hepatitis E 272, 274, 281
hepatitis E virus (HEV) 539–550

herpes simplex virus-1 293, 294, 298–300
herpesvirus capsid 491, 494, 497, 499, 501, 502, 506, 507
herpesvirus family 502
herpesvirus portal 499
higher-order packing of RNP 566, 567
HIV-1 175, 184–189
HLA class I 224
HLA class II 224
holey-carbon 300- or 400-mesh-Au grids (Quantifoil) 566
housekeeping and ambassadorial duties 555
HPV receptor 524–527
hsc70 609, 615
hsp70 609, 615
human immunodeficiency virus (HIV) 47, 52, 305, 328–331, 334–337, 339–341, 344–346, 348–351, 357, 358, 361, 362, 597, 598, 600, 601
human leukocyte antigen (HLA) 224, 225, 609, 617
human papillomavirus 517, 518, 523, 529, 530, 532
human rhinovirus 1–3, 5, 7
hVAP-33 159, 161, 162
hypervariable region (HVR) 582, 583, 586

icosahedral capsid 293, 300
in silico 234, 235, 238
indistinctive protein particles 561
influenza and rabies virus N protein monomers 558
Influenza A virus 46, 47
influenza virus 565, 567
insect cells 383, 386
integrin 63–72, 609, 617
interaction 7, 17, 21, 25
interferon 95–101, 106, 111

JC 609, 610, 621, 622

Kaposi's sarcoma-associated herpesvirus (KSHV) 485, 486
KKXX-like endoplasmic retention signal 582, 583

L1-capsomere 519–524
L2 cross-neutralization epitope 530, 532
L protein 554
L1 protein 517, 518, 519, 520, 522, 523, 527
L2 protein 517, 519, 520, 523–526, 533
L1 protein of the mammalian retrotransposon 564
late domain 318
lipid microdomains 43, 46, 49, 53, 55
lipid rafts 43, 44, 46, 51, 53, 55, 158–160, 317
lyssa-viruses 281, 556

M1 protein 48, 53, 54
M2 ion channel 50, 53, 54
M2 proteins 48
machine-learning 215, 219, 246
major histocompatibility complex (MHC) 224, 609, 616, 617, 619

major receptor group 3
malaria 597, 600, 601
Mamastrovirus 572
Marburg virus 311, 313, 317–319, 321,
mapping of oligomerization domains 558
Mason-Pfitzer monkey virus 184
mathematical model 210–212, 237
mathematical modeling 201, 210, 211
matrix protein 183, 312, 313, 316–318
membrane fusion 43, 51–53
MHC-binding peptides 224, 226
MHC class I 224, 226, 232, 238
MHC class II 224, 226, 227, 229
MHV 78–86
minichromosome 609, 610, 612–616, 620
minor receptor group 3
Molecular Interaction Network (MINet) 283, 286
moloney murine leukemia virus 176, 184
mouse leukemic virus 303–305
mouse mammary tumor virus 184
MS2 bacteriophage 174, 176
mucosal immunity 540, 546, 548, 549
multifunctional molecule 555
multiple epitope proteins (MEP) 203, 204

negative staining 558, 560, 565
negative-strand RNA viruses (NSRV) 553
nerve necrosis virus 449, 451
neuraminidase (NA) 48–51, 53, 54, 609, 617
NIAID 267–270, 272, 273
NMR spectroscopy 174, 177, 187, 188
N-N-interaction 557
N-oligomers 567
norovirus 381, 382, 389
Norwalk virus 381, 386–388
N protein 556, 561, 565
N-protein monomers 558, 565
N protein oligomers 553
N protein structure and functions 553
NS2 154
NS3 151–155
NS4A 154, 155
NS4B 154, 155, 159
NS5A 153–156, 159, 161
NS5B 151, 152, 154–157, 159, 161
NSRVs 555, 567
N-terminal domains of the monomers 559
nuclear localization signal (NLS) 582, 583, 585, 609, 613, 620
nucleocapsid 312, 313, 319, 320
nucleocapsid protein 183–185, 187–189, 553–562, 564–567
nucleoprotein (NP) 49, 312, 313, 315, 317, 319, 320
nucleotide triphosphatase 155

oligomers 556
open reading frame 153
oral vaccine 549

Ψ-site 185–189
packaging 173–178, 181, 184–190
panhandle 554, 561, 567
panhandle dissociation 561
panhandle-forming 561
panhandle-forming RNA sequences 567
panhandle-like structures 554
Pathogen Information Markup Language (PIML) 276, 283–285
PathPort 267, 268, 276, 282–285
PATRIC 267, 268, 270, 272–280, 282, 285
paramyxoviruses 52
PBCV-1 294–296
peptide design 233
physicochemical properties 211, 216, 220, 227, 231, 235, 240, 243
physicochemical scales 216
phytoreovirus 463, 464, 476
PI 609, 622
picornavirus 63, 64
plant viruses 293
polymers 553, 556
polyomavirus 609–612, 615, 617, 618, 620–623
portal vertex 410
prediction 203, 222, 223, 225–233, 241, 243–246
prediction of T-cell epitopes 203, 225
"prime-and-realign" mechanism 554
pRNA 406, 408, 410–415, 417–419, 421–424, 426–428, 432–436
procapsid 408, 410–415, 417–419, 421, 425, 426, 434
protease 78, 79, 82–86, 151, 153–155, 159
protein 200–212, 215, 216, 219–227, 233, 236–238, 244, 246
protein design 202, 207, 211, 223, 233, 236, 237, 246
protein engineering 200–203, 205, 207, 209–211, 237
pseudotype virus 125, 129–137
pseudovirion 517, 519, 523–533
putative RNA-binding domain 556
Puumala virus N protein 560

quantitative structure-activity-relationship (QSAR) 201, 202, 205–212, 223, 233, 235, 237

rabies and vesicular stomatitis viruses 566
rabies virus 559, 561, 565, 567
rabies virus nucleocapsid 565
random extensive sampling method 561
receptor 1–21, 23, 28–30, 125–131, 133, 135, 136, 139, 140
recombinant DNA 200, 208
recombinant Puumala virus 566
replication complex 151, 155, 157–162
replication of viral genome 554
replicon 152, 153, 157–159, 162
rhabdoviruses 556

ribonucleoprotein (RNP) 49, 50, 51, 54
ribosomal frameshifting signal (RFS) 573
Rice dwarf virus 464, 465
ring formation in the C-terminal region 561
ring-like structure 559, 567
RNA chaperone 556, 559, 561, 564, 567
RNA-dependent RNA polymerase 151, 155–157, 162
RNA encapsidation and replication 553, 556
RNA-panhandle 559, 561
RNA penetration 2, 23–25
RNA-RNA interactions 175, 189
RNA uncoating 2, 15, 27
RNP complexes 564
RNP of rabies virus 567
RNPs of vesicular stomatitis and rabies viruses 564
rubredoxin knuckle 184

SARS-CoV 77–86
scaffolding 487, 489, 491
Shine-Dalgarno sequence 176
signal transduction 112
signaling 67, 69, 72
single particle reconstruction 558, 560
Simian virus 609
SITUS 559, 562, 563
Sp1 609, 615
spike (S) protein 78
stoichiometry quantification 401–404, 436
structure 1, 4–8, 28, 30, 556, 557
subviral structure 289, 299, 300
SUMO-1-pathway members 556
supramolecule 463
surface glycoproteins (Gn and Gc) 554
SV40 609, 610–622

tail of the Gn protein 558
T-antigen 609, 615, 619
T-cell epitope 203, 223–225, 229–232
tegument proteins 484, 490–492, 494, 496–498, 502, 505, 507
terminal panhandle-forming regions 556
T helper cells 600
three dimensional homology 473, 476
ToolBus 276, 282, 283, 286
topoisomerase 564
trimer 464, 466, 470–480
tripartate genome 553
trypsin 572, 573, 575, 576, 578, 581, 589

uncoating 49, 53
unique binding region 556
uranyl acetate 560, 565

vaccine 328–330, 334, 337, 338, 340, 341, 345–347, 349, 350, 353, 358, 361, 362, 609, 623
vaccinia virus 289, 296, 297, 299, 300
vesicular stomatitis virus 561, 567
viral entry 80
viral envelope 83, 84
viral polymerase (L protein) 554

viral protein 200, 201, 207
viral RNA panhandle 556
viral RNA replication 555
viral structure 461
virus architecture 289
virus assembly 49, 50, 53, 553, 555
virus entry 13, 19, 30
virus-infected cells 290, 302, 303, 305
virus-like particles (VLPs) 205, 206, 313, 316, 317, 383, 384–386, 388–390, 392, 393, 539–550
virus-like particle vaccine 519
virus-receptor 7, 19
virus-specific 567
visualization of stereo pairs 562
VLP 609, 610, 618, 619, 622, 623
VMD 562, 563
VPg 573, 582, 584, 585, 586
VSV 129–133, 135

webservice 282
whirl-like 559

X-ray crystallography 174

zinc knuckle 184, 185, 188